Some Problems in Probability Theory Motivated by Number Theory and Mathematical Physics

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Yacine Barhoumi-Andréani

aus

Frankreich

Promotionskomitee

Prof. Dr. Michael Cranston (Irvine)
Prof. Dr. Ashkan Nikeghbali

Zürich, 2014

Pour Pierre Andréani (31 Mai 1922 - 12 Octobre 2013)

Contents

Summary

The aim of this thesis is the study of problems related to the theory of random matrices with a strong emphasis on their links with number theory. This is a direct application of the Keating-Snaith philosophy that consists in heuristically transposing problems from one world to another and deduce theorems or conjectures.

The first chapter is a survey of some recent developments in random matrix theory and its links with number theory, with some additional developments concerning the characteristic polynomial of a random unitary matrix.

The second chapter is an application of the Keating-Snaith philosophy in the case of a particular problem : the asymptotic number of zeros on the unit circle for linear combinations of characteristic polynomials of random unitary matrices, a problem initially motivated by number theory.

The third chapter studies mod-* convergence, a particular type of convergence that appears naturally in Number theory but which is less natural in Probability theory. When a Central Limit Theorem is involved in case of dependency, a correction to the independence is observed with mod-* convergence, in particular with random variables constructed from arithmetic considerations. We give a possible probabilistic interpretation of this convergence by means of a metrization, i.e. in terms of proximity to a "canonical" random variable that converges in this sense and we construct second order models that mimic the mod-* fluctuations in the case of the number of prime divisors of a random uniform integer (Selberg-Sathé theorems). We give moreover an explanation of the appearance of the corrective term for the arithmetic random variables involved in the Selberg-Sathé theorems by means of an additional randomisation, a natural operation of probability theory that allows to highlight hidden structures. The explanation is general to all structures sharing such a property and has a potential application to the moments conjecture.

The last chapter concerns the approximation of random variables converging in the mod-* sense, in particular since one probabilistic interpretation of the phenomenon can be done by means of a distance to a certain random variables, probabilistic methods of approximation become relevant, such as Stein's. An adaptation of Stein's method is done and amounts to a more accurate approximation in Kolmogorov distance, with potential applications to concrete problems such as Monte Carlo simulations.

Zusammenfassung

Das Ziel dieser Arbeit ist die Untersuchung von Problemen der Theorie der Zufallsmatrizen mit einem Schwerpunkt auf ihrer Verbindungen zu der Zahlentheorie. Dies ist eine direkte Anwendung der Keating-Snaith Philosophie, welche die Probleme heuristisch von einem Gebiet in das andere übersetzt und daraus Sätze und Vermutungen ableitet.

Das erste Kapitel ist ein Überblick über einige der jüngsten Entwicklungen in der Theorie der Zufallsmatrizen und ihrer Verbindungen zur Zahlentheorie, mit einigen zusätzlichen Entwicklungen bezüglich des charakteristisches Polynoms einer zufälligen unitären Matrix.

Das zweite Kapitel ist eine Anwendung der Keating-Snaith Philosophie auf eines besonderen Problems: die asymptotische Anzahl der Nullen auf dem Einheitskreis für Linearkombinationen von charakteristischen Polynomen von zufälligen unitären Matrizen, ein Problem zunächst durch Zahlentheorie motiviert.

Im dritten Kapitel wird mod-* Konvergenz untersucht, eine bestimmte Art von Konvergenz, die natürlich erscheint in der Zahlentheorie, aber in der Wahrscheinlichkeitstheorie weniger natürlich ist. Wenn in einem Model mit abhängigen Zufallsvariablen ein zentraler Grenzwertsatz beobachtet wird, ist eine Korrektur der Unabhängigkeit durch mod-* Konvergenz zu beobachten, insbesondere bei Zufallsgrössen, die aus arithmetischen Überlegungen konstruiert sind. Wir geben mögliche Wahrscheinlichkeitsinterpretation dieser Konvergenz durch eine Metrisierung, d.h. im Hinblick auf die Nähe zu einer "kanonischen" Zufallsgrösse, die in diesem Sinne konvergiert. Wir konstruieren Modelle zweiter Ordnung, die die mod-* Schwankungen imitieren am Beispiel der Anzahl der Primteiler uniform zufällig gewählten Zahl (Selberg-Sathe Sätze). Wir geben darüber hinaus eine Erklärung für die Erscheinung des Korrekturterms für die arithmetischen Zufallsgrössen in den Selberg-Sathe Theoremen durch eine zusätzliche Randomisierung. Dies ist ein natürlicher Vorgang in der Wahrscheinlichkeitstheorie, der beteiligte versteckte Strukturen hervorheben kann. Die Erklärung ist allgemein gültig für alle Strukturen, die eine solche Eigenschaft teilen und hat eine mögliche Anwendung bezüglich der Momentenvermutung.

Im letzten Kapitel geht es um die Angleichung der Zufallsgrössen, die im Mod-* Sinn konvergieren, insbesondere da eine probabilistische Interpretation des Phänomens über eine Entfernung zu gewissen Zufallsgrössen erfolgen kann, werden probabilistische Näherungsverfahren relevant, solche wie die Steinsche Methode. Eine Anpassung der Stein-Verfahren wird durchgeführt und erzeugt eine genauere Approximierung in Kolmogorov Abstand, mit möglichen Anwendungen auf konkrete Probleme wie die Monte-Carlo-Simulationen.

Remerciements

Oui, la rivière est profonde, et vastes et paisibles sont les eaux de mon enfance, dans un royaume que j'ai crû quitter il y a longtemps. Tous les chevaux du roi y pourraient boire ensemble à l'aise et tout leur saoul, sans les épuiser ! Elles viennent des glaciers, ardentes comme ces neiges lointaines, et elles ont la douceur de la glaise des plaines. Je viens de parler d'un de ces chevaux, qu'un enfant avait amené boire et qui a bu son content, longuement. Et j'en ai vu un autre venant boire un moment, sur les traces du même gamin si ça se trouve – mais là ça n'a pas traîné. Quelqu'un a dû le chasser. Et c'est tout, autant dire. Je vois pourtant des troupeaux innombrables de chevaux assoiffés qui errent dans la plaine – et pas plus tard que ce matin même leurs hennissements m'ont tiré du lit, à une heure indue, moi qui vais sur mes soixante ans et qui aime la tranquillité. Il n'y a rien eu à faire, il a fallu que je me lève. Ça me fait peine de les voir, à l'état de rosses efflanquées, alors que la bonne eau pourtant ne manque pas, ni les verts pâturages. Mais on dirait qu'un sortilège malveillant a été jeté sur cette contrée que j'avais connue accueillante, et condamné l'accès à ces eaux généreuses. Ou peut-être est-ce un coup monté par les maquignons du pays, pour faire tomber les prix qui sait ? Ou c'est un pays peut-être où il n'y a plus d'enfants pour mener boire les chevaux, où les chevaux ont soif, faute d'un gamin qui retrouve le chemin qui mène à la rivière...

A. Grothendieck,
Récoltes et semailles,
Promenade à travers une œuvre ou l'Enfant et la Mère.

Je voudrais avant tout remercier les personnes qui ont grandement contribué à mon éveil mathématique et à la réalisation de cette thèse. Ashkan Nikeghbali a été un directeur remarquable qui m'a introduit à de riches domaines de recherche, les matrices aléatoires et la convergence mod-*. Il m'a en outre donné la possibilité de le suivre dans ses conférences et écoles d'été, notamment en Chine et aux Etats-unis. Il m'a enfin permis de discuter en toute liberté avec d'autres scientifiques et d'échanger sur des sujets de recherche éloignés des siens, un premier pas vers l'autonomie qu'il a soutenu sans failles. Cette thèse et sa continuation lui doivent beaucoup.

Michael Cranston a eu la tâche, notoirement rebutante, de rapporter ce manuscript avec Ashkan et je leur suis redevable de nombreuses corrections. Mais Mike s'est avant tout déplacé spécialement des Etats-Unis pour la soutenance, ce qui mérite ma gratitude. Par la suite, nous avons eu des échanges mathématiques très intéressants qui, je l'espère, aboutiront dans un futur proche à de non moins intéressants théorèmes.

Certains des résultats de cette thèse ont été obtenus en collaboration avec Chris Hughes, Joseph Najnudel, Ashkan Nikeghbali et Réda Chhaibi. Je leur dois à tous mes remerciements pour m'avoir introduit à la recherche en collaboration et pour de fructueuses heures de discussions éreintantes. Je tiens par ailleurs à remercier tout particulièrement Joseph pour sa disponibilité et ses nombreuses explications, toujours lumineuses (ainsi que pour ses concerts de piano et son accompagnement de chants russes !). Sans oublier bien sûr Réda, ses baies, son tracteur et sa fonte (auxquels nous pourrions rajouter son amour de Brecht et du power yoga ainsi que sa conduite rapide et furieuse dans les rues de L.A.). Je tiens aussi à remercier Andrew Barbour pour ses nombreuses corrections, ses discussions lumineuses et ses encouragements quant à la dernière partie de cette thèse.

La possibilité d'un doctorat a été initiée de façon décisive par Colin Umansky et ses passionnantes questions concernant le mouvement brownien. Qu'il soit ici remercié de tous ses efforts.

Cette thèse n'aurait pas été possible sans l'excellent enseignement du master Processus Stochastiques de l'UPMC. Je tiens à remercier tous les professeurs dont j'ai suivi les cours en 2008-2009 : Jean Bertoin, Philippe Biane, Jean Jacod, Silvie Méléart, Zhan Shi, Marc Yor et Lorenzo Zambotti. Quatre personnes sont à remercier tout particulièrement : Jean Jacod pour m'avoir permis de suivre le master, Marc Yor pour son incroyable cours de temps locaux et excursions ainsi que pour ses conseils et recommandations dans le choix de mon directeur, et Philippe Biane et Lorenzo Zambotti pour leurs recommandations. Par la suite, j'ai bénéficié des excellents cours d'Alain-Sol Sznitman, Kurt Johannsson, Thomas Kappeler, Emmanuel Kowalski, Pierre-Loïc Méliot et Paul-Olivier Dehaye que je remercie également.

Il m'est aussi arrivé d'embêter de nombreuses personnes sur des questions mathématiques parfois pointilleuses, et je les remercie pour leurs réponses : Andrew Barbour, Djalil Chafai, Patrik Ferrari, Yan Fyodorov, Thomas Kappeler, Jon Keating, Camillo de Lellis, Simon Pépin-Lehalleur, Todor Popov et Pedro Vaz.

Mes pérégrinations mathématiques auront été l'occasion de rencontres et d'amitiés, mathématiques ou non (ne rejettons pas nos amis les physiciens). Celle avant tout d'Olivier Hénard, qui n'a pas son pareil pour dessiner des arbres en forme de bulles ou de soucoupes volantes ; malgré tous ses efforts pédagogiques, je reste imperméable à l'art abstrait de l'élagage, mais ce n'est pas sans espoir (qui vivra verra). Pascal "Brutal" Maillard et sa Pauline Brutale, grands brancheurs (branchés ?) devant l'éternel m'ont gracieusement fait squatter leur appartement et introduit à la staritude parisienne de leur voisin star de cinoche (Mathieu 4ever). Simon Pépin-Lehalleur mérite l'award du voisin de bureau le plus agréable du monde, et je tiens aussi à saluer son incroyable culture mathématique et extra-mathématique, notamment en psychologie, sociologie, anthropologie, religion, natation, etc. Eh oui, Simon, c'est Monsieur Plus ! L'award pourrait aussi être décerné à mon autre voisin de bureau Steffen Weil et son amour de *Sucker Punch*. Et n'oublions pas Gaétan Borot et son appartement de glandu qui m'a souvent accueilli, Pierre-Loïc Méliot, grand fan des *MC Warriors*, Todor Popov et ses soirées à Sofia, Ivan Dornic et sa cabane au fond du jardin dans la Corse bien paumée, Miguel Tierz et sa grande amitié avec les se(r)vices secrets israéliens, tout particulièrement dans les aéroports, Andrew Kresh, sa voix de Stentor (Russe !) et son amour du cinéma zurichois (the *delay effect*), et bien d'autres que j'oublie sûrement.

Je remercie par ailleurs mes amis et collègues de l'université : Dirk Zeindler, Kenneth Maples, Julia Storm, Hilmar Mai, Martin Wahl, Doerte Kreher, Félix Rubin, Kim Dang, Alexandre Usnich, Akiko Yasumoto, Atsushi Aoyama, Ren Long Miao, Marko Zivkovic, Sara Filippini, Valentin Féray, Joseph Ayoub, Alessandra Cipriani, Erich Baur, Leif Doering, Pierre Nolin, Steven Geinitz, Tamara Widmer, Hasan Inci, Yannick Widmer, Ivan et Eva Contreras, Utsav Choudhury, Cyril Marzouk, David Belius, Pierre-François Rodriguez, Rajat Hazra, Luca Avena, Amel Bentata, Anne Feidt, Peter Bruin, Elisabetta Chiodaroli, Christoph Luchsinger, Damien Simon, Andreas Kyprianou, Hung Bui, Nicolas Robles, Victoria Hoskins, Thomas

Preu, Asieh Parsania, Sara Daneri et Karen Yessarian (là, j'espère que je n'en ai pas oublié...).

Je remercie également mes amis : Julien Gaudremeau, Antoine Chopineau, Guillaume Mitrecey, Julien Fracque, Jonathan Brossard, Emmanuel Decoux, Maud Lucq, Thibaut de Vergnette, Mohamed Annouch, Hani Ouragini, Julien Loknar, Yoann Rabatel, Frédéric Khamsing, Mathieu Ducros, Thomas Hoppe, David Ginsbourger, Guillaume Pirot, Philippe Sourlas, Laura Zaugg, Min Cui, Yves Laffiché et Shohei Uesugi.

Je remercie enfin les secrétaires d'Irchel : Myriam Frank, Nha Nguyen, Carola Herren, Gerda Schacher, Bettina Kurth, Claudia Burkhardt et Carla Allenbach, ainsi que l'équipe IT composée de Carsten Rose et Rafael Ostertag.

Le mot de la fin sera pour ma famille, mon frère Karim, ma mère Parsilie ainsi que ma grand-mère Jeanne et mon grand-père Pierre qui n'aura pas eu la nouvelle de la soutenance et à qui cette thèse est dédiée.

Notations

$\mathbb{P}_{\mathcal{U}(\mathfrak{S}_n)}$ Uniform measure on $\mathfrak{S}_n$

$\mathbb{P}_\theta^{(n)}$ Ewens measure on $\mathfrak{S}_n$

$\mathbb{M}_q^{(n)}$ Mallows measure on $\mathfrak{S}_n$

$[n]_q$ q-notation $1 + q + \cdots + q^{n-1}$

$\mathcal{O}_n(\mathbb{R})$ Orthogonal group

$\mathcal{U}_n(\mathbb{C})$ Unitary group

P_σ Geometric representation $P_\sigma = (\delta_{i,\sigma(j)})_{i,j\leqslant n}$

$\mathbb{P}_{O(n)}, \mathbb{P}_{U(n)}$ Haar measures of $\mathcal{O}_n(\mathbb{R})$ and $\mathcal{U}_n(\mathbb{C})$

$COE(n)$ Circular Orthogonal Ensemble

$CUE(n)$ Circular Unitary Ensemble

$h_{u,v}$ Householder matrices

H_u Householder matrices $h_{e_1,u}$

$\mathcal{C}(\sigma), \mathrm{ct}(\sigma)$ Cycle structure and cycle type of $\sigma \in \mathfrak{S}_n$

$\mathscr{P}(\theta)$ Poisson distribution of parameter θ

$\mathscr{P}(\theta)$ Poisson distribution of parameter θ

z_λ Number of equivalence classes of type λ

$\ell(\lambda)$ Length of $\lambda \vdash n$

$\Delta(z_1, \ldots, z_n)$ Vandermonde determinant

$[\![1, n]\!]$ Set $\{1, 2, \ldots, n\}$

d_λ Dimension of the irreducible module V^λ of $\mathfrak{S}_n$

$e_k(z_1, \ldots, z_n)$ k-th elementary symmetric polynomial

$\Phi_U(z)$ Characteristic polynomial of U in $z \in \mathbb{C}$

$Z_U(\theta)$ Characteristic polynomial of U on the unit circle, for $\theta \in [0, 2\pi[$

$\zeta(s)$ Riemann Zeta function

$X_n \xrightarrow[n \to +\infty]{\mathcal{L}} X$ Convergence in law / in distribution of $(X_n)_n$

$\mathscr{U}(A)$ Uniform distribution on A

$A(\lambda), M(\lambda)$ Arithmetic and Matrix factors

$\mathcal{G}(z), G(z)$ Barnes G-function

s_λ Schur function of label $\lambda \vdash n$

$sc_k^{(n)}$ Secular coefficient of order k of a matrix of size n

$K_0(x)$ Modified Bessel function of second kind

$\|f\|_p, \|f\|_\infty$ L^p and L^∞ norms of f

$SST(\lambda)$ Set of semi-standard Young tableaux of shape λ

$SU(N)$ Group of Special Unitary matrices of size N

$\mathbb{P}_{U(N)}, \mathbb{P}_{SU(N)}$ Normalised Haar measures on these sets

λ, λ_α Lebesgue measure on $\mathbb{R}$, resp. normalised Lebesgue measure on an interval of length α

$|X|$ Number of elements of the finite set X

$\mathbb{P}_{SU(N)}^{(n)}, \mathbb{E}_{SU(N)}^{(n)}$ n-fold product of the Haar measure on $SU(N)$ and corresponding expectation

I_ν Bessel function of first type

$\mathscr{C}(\alpha)$ Symmetric Cauchy distribution of parameter α

$\mathrm{Gb}(1)$ Gumbel distribution of parameter 1

$\omega(N)$ Number of prime divisors of $N \in \mathbb{N}$

$H_n^{(\mathcal{P})}$ Prime harmonic sum

$\Phi_A(z), \Phi_U(z)$ Arithmetic and Matrix factor

$Z \bullet \mathbb{P}_X$ Bias of the measure $\mathbb{P}_X$ with the weight Z

$X^{(s)}$ Size-bias of X

$(Z_n, \gamma_n)_n \xrightarrow[n \to +\infty]{\text{mod-P}} \Phi$ Mod-Poisson convergence of $(Z_n)_n$ at speed $(\gamma_n)_n$

$(Z_n, \gamma_n)_n \xrightarrow[n \to +\infty]{\text{mod-G}} \Phi$ Mod-Gaussian convergence of $(Z_n)_n$ at speed $(\gamma_n)_n$

$\mathrm{CIP}(G)$ Cycle Index Polynomial of a group G

$(v_p)_{p\in\mathcal{P}}$ Valuation structure

$\boldsymbol{\delta\zeta}(\alpha)$ Delta-Zeta distribution of parameter $\alpha > 1$

$\mathrm{Li}_\alpha(z)$ Polylogarithm function

$\zeta_{\mathcal{P}}(z)$ Prime Zeta function

d_{Kol} Kolmogorov distance

d_{TV} Total variation distance

W_1 Wasserstein distance

$X^{(0)}$ Zero-bias transform of the random variable X

$d_{\mathcal{H}_\Phi}$ Probabilistic distance on the set $\mathcal{H}_\Phi$

Chapter 1

Introduction

This Ph.D. thesis finds its motivations in problems occuring at the interface of probability theory, number theory and mathematical physics. Over the past two decades, there have been many new results at the interface of random matrix theory and analytic number theory that can be considered as evidence for the zeros of the Riemann zeta function being statistically distributed as eigenvalues of large random matrices (GUE matrices or Haar distributed unitary matrices). The references [74], [61] and [86] give a detailed account with many references (see also [60] for the function field framework).

Since the seminal papers by Keating and Snaith [62, 63], it is believed that the characteristic polynomial of random unitary matrices on the unit circle models very accurately the value distribution of the Riemann zeta function or more generally L-functions on the critical line. This analogy was used by Keating and Snaith to produce the moments conjecture and since then the characteristic polynomial has been the topic of many research papers, and the moments of the characteristic polynomial have now been derived with many different methods, e.g. representation theoretic methods (see [22, 33, 81]), super-symmetry method (see [74]), analytic methods (Toeplitz determinant methods as explained in the lecture by E. Basor in [74], orthogonal polynomials on the unit circle method [65]) or probabilistic methods ([19]), each method bringing a new insight to the problem. Many more fine properties of the characteristic polynomial have been established, e.g. large deviations principle in [49], local limit theorems in [67], the analogue of the moments conjecture for finite field zeta functions [53], etc. Moreover, thanks to this analogy, one has been able to perform calculations in the random matrix world whose analogue in the number theory world seems currently out of reach and to produce conjectures for the analogue arithmetic objects (see [90] for a recent account).

This introduction which is partly inspired by [30, 80] summarizes some recent developments about the characteristic polynomial of a random unitary matrices on the unit circle, including personal developments (theorems 1.4.9 and 1.4.13). The point of view to expose the motivations of the study is not historical, the choice was made to stress similarities with other algebraic objects such as e.g. random permutations.

1.1 From random permutations to random isometries

1.1.1 Simulating the uniform measure on the symmetric group

There are natural links between permutations and matrices. For a permutation $\sigma \in \mathfrak{S}_n$, one can consider the permutation matrix $P_\sigma := \left(\mathbb{1}_{\{i=\sigma(j)\}}\right)_{1\leqslant i,j\leqslant n}$ or more generally a group morphism from $\mathfrak{S}_n$ to a space of matrices, that is, a representation of $\mathfrak{S}_n$. From this point of view, a random permutation induces a random matrix. It is important to notice that in the case of the representation $P : \sigma \mapsto P_\sigma$, we get an orthogonal matrix since $P_\sigma^{-1} = P_{\sigma^{-1}} = P_\sigma^*$.

There are several ways to define a measure on $\mathfrak{S}_n$. The most famous one is the uniform measure defined for all $\sigma \in \mathfrak{S}_n$ by

$$\mathbb{P}_{\mathcal{U}(\mathfrak{S}_n)}(\sigma) := \frac{1}{n!} \tag{1.1}$$

For simulation purposes, one can ask the question of selecting at random such a uniform permutation, what we note $\sigma \sim \mathcal{U}(\mathfrak{S}_n)$. Of course, a basic algorithm of complexity $O(n!)$ consists in generating all the permutations of $\mathfrak{S}_n$, numeroting them, and selecting randomly a number between 1 and $n!$. But a more refined algorithm exists and can be stated as follows : start with the identity permutation id_n and for $k \in [\![2, n]\!]$, select $U_k \sim \mathcal{U}\left([\![1, k]\!]\right)$, and exchange k and U_k in the word of the permutation.

More formally, such a permutation is obtained via the recursive formula :

$$\begin{cases} \sigma_0 & = id_n \\ \sigma_{k+1} & = \sigma_k\,(k+1, U_{k+1}) \end{cases} \tag{1.2}$$

where (i, j) is the permutation that exchanges i and j if $i \neq j$ and is understood as the identity if $i = j$, and $(U_k)_k$ is a sequence of independent random variables uniformly distributed as before.

This algorithm induces a way to label the permutations by means of a "factorial alphabet" (see [69, 32]). It is enough to simulate n uniform random variables and to multiply n permutations, which gives a fast way of simulating $\sigma_n \sim \mathcal{U}(\mathfrak{S}_n)$. The fact that the permutation obtained at the end of the process is uniform on $\mathfrak{S}_n$ is reminiscent of the following identity in the group algebra $\mathbb{C}\left[\mathfrak{S}_n\right]$

$$\frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\sigma = \prod_{k=1}^{n}\left(\frac{1}{k}\sum_{\ell=1}^{k}(\ell, k)\right) \tag{1.3}$$

Here, we have considered that a measure on $\mathfrak{S}_n$ is an element $\mathbb{P} := \sum_{\sigma\in\mathfrak{S}_n}\mathbb{P}(\sigma)\sigma \in \mathbb{C}\left[\mathfrak{S}_n\right]$ that satisfies $\mathbb{P}(\sigma) \geqslant 0$ for all $\sigma \in \mathfrak{S}_n$ and $\sum_{\sigma\in\mathfrak{S}_n}\mathbb{P}(\sigma) = 1$, and that the non-commutative product is taken increasingly (we will adopt this convention from now). With these conventions, the last identity expresses the disintegration of the uniform measure on $\mathfrak{S}_n$ in terms of a product of independent measures on the cosets $\mathfrak{S}_{k+1}/\mathfrak{S}_k$ once the choice of the coset representatives is fixed, in this case by setting

$$\mathfrak{S}_{k+1}/\mathfrak{S}_k = \{(\ell, k+1)\mathfrak{S}_k, \ell \in [\![1, k+1]\!]\}$$

Another choice of coset representatives would induce the same exact disintegration, i.e. if

$$\mathfrak{S}_{k+1}/\mathfrak{S}_k = \left\{R_\ell^{(k+1)}\mathfrak{S}_k, \ell \in [\![1, k+1]\!]\right\}$$

then

$$\frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\sigma=\prod_{k=1}^{n}\left(\frac{1}{k}\sum_{\ell=1}^{k}R_\ell^{(k)}\right)$$

In a more intuitive way, the coset representative tells us how to "place" $k+1$ in the permutation σ_k constructed with (1.2) hence belonging to $\mathfrak{S}_k$ via the natural embedding $\mathfrak{S}_k \hookrightarrow \mathfrak{S}_{k+1}$ that consists in considering $k+1$ as a fixed point of a permutation of $\mathfrak{S}_k$. In the case of the system $\{(i,j)\}_{i,j}$, an interpretation in terms of the cycle structure of the permutation can be given, as seen in what follows.

1.1.2 Some distinguished measures on the symmetric group

The Ewens measure

The uniform measure is not the only interesting measure that enjoys a disintegration in terms of product of independent random coset representatives. The general class of such measures could be defined by specifying the choice of the cosets representatives and the measure that we associate with them. A particular case of interest is given by the *Ewens measures* (see for instance [2]) defined by

$$\mathbb{P}_\theta^{(n)}(\sigma):=\frac{\theta^{C(\sigma)}}{\theta(\theta+1)\dots(\theta+n-1)} \tag{1.4}$$

where $\theta\geqslant 0$ and $C(\sigma)$ designates the number of cycles of σ.

We can remark that this measure is equal to the uniform measure $\mathbb{P}_{\mathcal{U}(\mathfrak{S}_n)}$ for $\theta=1$ and to the Dirac measure in the identity for $\theta=+\infty$. For $\theta=0$, this is the uniform measure on the cyclic permutations of size n. The disintegration on the coset representatives is given by

$$\frac{1}{\theta(\theta+1)\dots(\theta+n-1)}\sum_{\sigma\in\mathfrak{S}_n}\theta^{C(\sigma)}\sigma=\prod_{k=1}^{n}\left(\frac{1}{\theta+k-1}\left(\theta\, id_n+\sum_{\ell=1}^{k-1}(\ell,k)\right)\right) \tag{1.5}$$

wich consists in a bias of the uniform distribution on $[\![1,k]\!]$ by imposing a probability proportional to θ to choose k, i.e. if $\sigma_n\sim\mathbb{P}_\theta^{(n)}$, then

$$\sigma_n=\left(1,X_1^{(\theta)}\right)\left(2,X_2^{(\theta)}\right)\dots\left(n,X_n^{(\theta)}\right)$$

where $X_k^{(\theta)}\in[\![1,k]\!]$ and for all $\ell\in[\![1,k]\!]$

$$\mathbb{P}\left(X_k^{(\theta)}=\ell\right)=\frac{\theta}{\theta+k-1}\mathbb{1}_{\{\ell=k\}}+\frac{1}{\theta+k-1}\mathbb{1}_{\{\ell\neq k\}}=\frac{\theta^{\mathbb{1}_{\{\ell=k\}}}}{\theta+k-1}$$

The intuitive way of seeing the operation of composing to the right by $(\ell,k+1)$ a permutation σ_k such that $\sigma_k(k+1)=k+1$ consists in placing $k+1$ next to ℓ in the cycle writing of the permutation. For example, if $\sigma=(136)(245)$ and $k=3$, we have $\sigma(3,7)=(1376)(245)$ and we have positioned 7 just after $k=3$ in its cyclic writing. One can draw a pictural representation of such a cycle structure by means of "tables" where the numbers sit. This way of constructing a random Ewens permutation hence translates into placing $k+1$ to a new table with probability $\theta/(\theta+k-1)$ and to place it to an existing table with probability $1/(\theta+k-1)$: this is the so-called *chinese restaurant process* (see [2]).

Remark 1.1.1. The equality (1.5) is clearly equivalent to the equality

$$\sum_{\sigma\in\mathfrak{S}_n} \theta^{C(\sigma)}\sigma = \prod_{k=1}^{n}\left(\theta\, id_n + \sum_{\ell=1}^{k-1}(\ell,k)\right)$$

which is known in algebra as the *Jucys-Murphy identity* (see [57, 77]). Indeed, if we define the Jucys-Murphy elements $(J_k)_k$ by $J_1 = 0$, $J_2 = (1,2)$ and $J_k = \sum_{i=1}^{k-1}(i,k)$ for all $k \geqslant 3$, the equality (1.5) rewrites into $\sum_{\sigma\in\mathfrak{S}_n} \theta^{C(\sigma)}\sigma = \prod_{k=1}^{n}(\theta + J_k)$.

The Mallows measure

Another interesting measure that enjoys the same type of disintegration in terms of product of independent random coset representatives as the Ewens measure is the following *Mallows measure* (see [72])

$$\mathbb{M}_q^{(n)}(\sigma) := \frac{q^{\mathrm{inv}(\sigma)}}{(1+q)(1+q+q^2)\dots(1+q+\dots+q^{n-1})} \tag{1.6}$$

where $q \geqslant 0$ and $\mathrm{inv}(\sigma)$ designates the number of inversions of σ defined by

$$\mathrm{inv}(\sigma) := \sum_{1\leqslant i<j\leqslant n} \mathbb{1}_{\{\sigma(i)>\sigma(j)\}} \tag{1.7}$$

As in the case of the Ewens measure, all the Mallows measures can be coupled into a process $(\tau_n)_n$ of one-dimensional marginals $\tau_n \sim \mathbb{M}_q^{(n)}$. But the system of coset representatives is the one of the *word writing*. A permutation can be written as a product of disjoint cycles or as a word. For example,

$$\begin{aligned}\sigma &= (136)(245) \qquad \text{(cycles)}\\ &= \langle 346521\rangle \qquad \text{(word)}\end{aligned}$$

Define

$$[k, n+1] := (n+1, n, \dots, k+1, k) = \langle 1\ 2 \cdots k{-}1\ n{+}1\ k \cdots n\rangle$$

with the convention that $[n+1, n+1] := id$.

The intuitive way of seeing the operation of composing to the right by $[\ell, k+1]$ a permutation σ_k such that $\sigma_k(k+1) = k+1$ consists in placing $k+1$ next to ℓ in the word writing of the permutation. For example, if $\sigma = (136)(245) = \langle 346521\rangle$ and $k = 3$, we have $\sigma\,[3,7] = \langle 3476521\rangle$ which amounts to position 7 as the third symbol of the writing as a word. Here, the relevant pictural representation for such a word structure is just a line, the coordinates $(\sigma(1),\dots,\sigma(n))$ being "particules" on the line. The process $(\tau_n)_n$ consists then in the insertion of a new particule.

The definition of this process is the following : if $\tau_n \sim \mathbb{M}_q^{(n)}$, to construct $\tau_{n+1} \sim \mathbb{M}_q^{(n+1)}$ using τ_n, we place $n+1$ in the word of τ_n (as $\tau_n(n+1) = n+1$). To do so, we define the geometric distribution on $[\![1,n]\!]$ by

$$\mu_q^{(n)} := \frac{1}{[n]_q}\left(\sum_{k=1}^{n} q^{n-k}\delta_k\right)$$

where $[n]_q := 1 + q + \cdots + q^{n-1}$. Let $Y_q^{(n)} \sim \mu_q^{(n)}$ be choosen independently of τ_{n-1} and set

$$\tau_n = \left[1, Y_q^{(1)}\right]\left[2, Y_q^{(2)}\right]\dots\left[n, Y_q^{(n)}\right] = \tau_{n-1}\left[n, Y_q^{(n)}\right]$$

Then, we have

$$\mathbb{M}_q^{(n+1)} = \mathbb{M}_q^{(n)}\left(\frac{1}{[k+1]_q}\sum_{\ell=1}^{k} q^{n-\ell}\,[\ell, k]\right) \tag{1.8}$$

which is equivalent to the formula

$$\sum_{\sigma\in\mathfrak{S}_n} q^{\mathrm{inv}(\sigma)}\sigma = \prod_{k=1}^{n}\left(\sum_{\ell=1}^{k} q^{n-\ell}\,[\ell, k]\right) \tag{1.9}$$

1.1.3 Coupling all the symmetric groups

The latest constructions of the measures on $\mathfrak{S}_n$ share the same property : by the embedding $\mathfrak{S}_n \hookrightarrow \mathfrak{S}_{n+1}$, it can be extended to all the symmetric groups at the same time, and one can define a probability measure on the projective limit of the symmetric groups, this projective limit being taken with respect to the projections associated with the coset representatives, that is, for a system of coset representatives $\left\{R_\ell^{(k)}, \ell \leqslant k, k \geqslant 1\right\}$, with respect to the projections

$$p_{n+k\to n} : \sigma_{n+k} := \prod_{\ell=1}^{n+k} R_{s_\ell}^{(\ell)} \in \mathfrak{S}_{n+k} \mapsto \sigma_n := \prod_{\ell=1}^{n} R_{s_\ell}^{(\ell)} \in \mathfrak{S}_n$$

This is the sequence $(s_1, s_2, \dots) \in \prod_{k\geqslant 1}[\![1, k]\!]$ that defines the limiting permutation $\sigma_\infty \in \varprojlim_n(\mathfrak{S}_n, p_{n+1\to n})$ and defining a probability measure on this space is equivalent to defining a probability measure on $\prod_{k\geqslant 1}[\![1, k]\!]$.

In the case of the Ewens measure or the Mallows measure, $(s_1, s_2, \dots)$ are in addition independent random variables, and the process $(\sigma_n)_n$ is markovian. This allows to define the Ewens measure (resp. the Mallows measure) on the projective limit $\mathfrak{S}_{(\infty)}$ associated to the cyclic system of coset representative (resp. $\mathfrak{S}_{\langle\infty\rangle}$ with the word system). Note that in both cases, the limit space is no more a group since the projections are not group morphisms, and the projective limit is taken in the category of sets : $\mathfrak{S}_{(\infty)}$ and $\mathfrak{S}_{\langle\infty\rangle}$ are only sets. This restriction makes it hard to consider a Haar measure on $\mathfrak{S}_{(\infty)}$ and $\mathfrak{S}_{\langle\infty\rangle}$, but a substitute of Haar measure is precisely played by $\mathbb{P}_1^{(\infty)}$ and $\mathbb{M}_1^{(\infty)}$ in the action of the injective limit $\varinjlim_n \mathfrak{S}_n$.

The space $\mathfrak{S}_{(\infty)}$ was defined by Kerov as the space of *virtual permutations* (see [64]). It is a compact topological space for the projective limit topology and any projective family of probability distributions (that is, sequences $(\mathbb{P}_n)_n$ of probabilities that are consistent with the projections in the sense that $\mathbb{P}_{n+1} \circ p_{n+1\to n}^{-1} = \mathbb{P}_n$) defines a probability on $\mathfrak{S}_{(\infty)}$ by Kolmogorov's theorem. References for the space $\mathfrak{S}_{\langle\infty\rangle}$ that we could call *space of virtual words* can be found in [43].

As we will see, a replica of these constructions can be done for matrix groups that will allow to couple all the dimensions of the groups.

1.2 Random isometries

1.2.1 Some ensembles of random matrices

The spaces of orthogonal and unitary matrices can be tought of as the most direct generalisation of the symmetric group. If we denote these groups by

$$\mathcal{O}_n(\mathbb{R}) := \left\{M \in \mathcal{M}_n(\mathbb{R}) \;/\; {}^tMM = I_n\right\}$$
$$\mathcal{U}_n(\mathbb{C}) := \left\{M \in \mathcal{M}_n(\mathbb{C}) \;/\; M^*M = I_n\right\}$$

we have a natural embedding of $\mathfrak{S}_n$ in $\mathcal{O}_n(\mathbb{R})$ and $\mathcal{U}_n(\mathbb{C})$ given by the geometric representation $P : \sigma \in \mathfrak{S}_n \mapsto P_\sigma = (\delta_{i,\sigma(j)})_{i,j\leqslant n}$.

These groups are moreover compact and their representation theory can be done in the same maneer as for the symmetric group by replacing the discrete average on $\mathfrak{S}_n$ by the average with respect to their normalised Haar measure (see e.g. [102]). We recall the

Definition 1.2.1 (Haar measure)**.** Let G be a group. Its left (resp. right) Haar measure is a measure that is translation invariant by the action of G on the left (resp. right), that is, for all $g \in G$, for all measurable set $A \subset G$, if λ designates such a measure,

$$\lambda(gA) = \lambda(A)$$

One can show that if left and right Haar measures exist, they coincide and are uniquely determined up to a scaling factor ; in particular, if the group is compact, one can fix this factor by imposing a probability measure. We speak of normalized Haar measure in such a case. From now on, we will only consider the case of a normalised Haar measure for all compact groups encoutered. In particular, if U is a random matrix selected according to the Haar measure of $\mathcal{O}_n(\mathbb{R})$ (resp. $\mathcal{U}_n(\mathbb{C})$), we will denote it by $U \sim \mathrm{Haar}\left(\mathcal{O}_n(\mathbb{R})\right)$ (resp. $U \sim \mathrm{Haar}\left(\mathcal{U}_n(\mathbb{C})\right)$).

The Haar measures of $\mathcal{O}_n(\mathbb{R})$ and $\mathcal{U}_n(\mathbb{C})$ will be denoted by $\mathbb{P}_{O(n)}$ and $\mathbb{P}_{U(n)}$. The couples $(\mathcal{O}_n(\mathbb{R}), \mathbb{P}_{O(n)})$ and $(\mathcal{U}_n(\mathbb{C}), \mathbb{P}_{U(n)})$ are said respectively Circular Orthogonal Ensembles and Circular Unitary Ensembles of size n and are denoted by $COE(n)$ and $CUE(n)$.

In the same vein as the symmetric group, one can ask how to generate efficiently a random orthogonal matrix selected according to the Haar measure of $\mathcal{U}_n(\mathbb{C})$ (resp. $\mathcal{O}_n(\mathbb{R})$), and in particular, if the coset disintegration (1.3) still holds. A first glimpse at the structure of these groups gives

$$\mathcal{O}_{n+1}(\mathbb{R})/\mathcal{O}_n(\mathbb{R}) \simeq \mathbb{S}^n(\mathbb{R}) := \left\{x \in \mathbb{R}^{n+1} \;/\; |x| = 1\right\}$$
$$\mathcal{U}_{n+1}(\mathbb{C})/\mathcal{U}_n(\mathbb{C}) \simeq \mathbb{S}^n(\mathbb{C}) := \left\{x \in \mathbb{C}^{n+1} \;/\; |x| = 1\right\}$$

Hence, one can expect a similar disintegration by means of the uniform measure on $\mathbb{S}^n$ once specified a choice of coset representatives.

In the embedding $\mathfrak{S}_n \hookrightarrow \mathcal{O}_n(\mathbb{R})$, a natural way to embed the generators (i,j) is to consider the Householder matrices $P_{(i,j)} := h_{i,j} := I_n - (e_i - e_j)\cdot {}^t(e_i - e_j)$ with respect to a given orthonormal basis $(e_1, \dots, e_n)$ of $\mathbb{R}^n$. Such matrices satisfy $h_{i,j}(e_i) = e_j$, $h_{i,j}(e_j) = e_i$ and

$h_{i,j}(e_k) = e_k$ for all $k \neq i, j$. A natural generalisation of these matrices in $\mathcal{O}_n(\mathbb{R})$ is thus the Householder matrices

$$h_{u,v} := I_n - 2\frac{(u-v)\cdot {}^t(u-v)}{|u-v|^2}$$

that exchange the different normalised vectors u and v (i.e. $|u-v| \neq 0$ and $|u| = |v| = 1$) and that fixes the orthogonal of $\mathrm{span}(u-v)$. It is clear that such matrices are orthogonal. Hence, the matrices

$$H_u := h_{e_1,u} = I_n - \frac{(u-e_1)\,{}^t(u-e_1)}{1 - {}^te_1 u}$$

are parametrized by $u \in \mathbb{S}^{n-1}\backslash\{e_1\}$ and form a system of coset representatives of $\mathcal{O}_n(\mathbb{R})/\mathcal{O}_{n-1}(\mathbb{R})$. Of course, the embedding $\mathcal{O}_n(\mathbb{R}) \hookrightarrow \mathcal{O}_{n+1}(\mathbb{R})$ is given by $M \mapsto 1 \oplus M = \begin{pmatrix} 1 & 0 \\ 0 & M \end{pmatrix}$.

A natural possible generalisation of (1.3) follows

Theorem 1.2.2 (Disintegration of the Haar measure of $\mathcal{O}_n(\mathbb{R})$, [32, 73])**.** *Let $(u_k)_{k\leqslant n}$ be independent random unit vectors such that for all $k \leqslant n$, $u_k \sim \mathcal{U}\left(\mathbb{S}^{k-1}\right)$. Then, if $M \sim$* Haar $(\mathcal{O}_n(\mathbb{R}))$, *we have the equality in law*

$$M \overset{\mathcal{L}}{=} H_{u_1} H_{u_2} \dots H_{u_n}$$

Note that another choice of coset representatives would give the exact same result. Note also that the same theorem holds for $\mathcal{U}_n(\mathbb{C})$ with slight changes of definition : the Householder matrices here described are not anymore orthogonal due to the complex scalar product. Indeed, if one defines

$$H_u := h_{e_1,u} = I_n - 2\frac{(u-e_1)\,(u-e_1)^*}{|e_1-u|^2} = I_n - \frac{(u-e_1)\,(u-e_1)^*}{1 - \mathfrak{Re}\,(e_1^* u)}$$

where $u^* = {}^t\bar{u}$ is the adjoint of u, one has

$$H_u(u) = u - \frac{(u-e_1)\,(u-e_1)^* u}{1 - \mathfrak{Re}\,(e_1^* u)} = u - \frac{1 - e_1^* u}{1 - \mathfrak{Re}\,(e_1^* u)}(u - e_1) \neq e_1$$

The problem comes thus from the real part in the scalar product and the disintegration theorem 1.2.2 has to be modified in consequence to get an analogous result. One way to proceed is to consider complex proper reflections, that is, norm preserving automorphisms of $\mathbb{C}^n$ that leave exactly one hyperplane fixed and that can be written as

$$s_{a,\lambda} : x \mapsto= x - (1-\lambda)\frac{a^* x}{|a|^2} a$$

with $a \in \mathbb{C}^n$ and $\lambda \in \mathbb{C}$ with $|\lambda| = 1$. If $u \neq e_1$, such a reflection mapping e_1 onto u is obtained by taking $a = e_1 - u$ and $\lambda = -(1 - e_1^* u)/(1 - \overline{e_1^* u})$. We thus set in this case

$$H_u : x \mapsto x - \left(1 + \frac{1 - e_1^* u}{1 - \overline{e_1^* u}}\right) \frac{(e_1 - u)^* x}{|e_1 - u|^2}(e_1 - u) \tag{1.10}$$

We leave the reader to [19, 13, 20] for further generalisations and other groups.

1.2.2 Coupling all the unitary groups

As for the case of the symmetric group, one can define the projections

$$\pi_{n+1\to n} : \begin{pmatrix} \mathcal{U}_{n+1}(\mathbb{C}) & \longrightarrow & \mathcal{U}_n(\mathbb{C}) \\ \prod_{k=1}^{n+1} H_{u_k} & \mapsto & \prod_{k=1}^{n} H_{u_k} \end{pmatrix}$$

The push-forward of $\mathbb{P}_{U(n+1)}$ under $\pi_{n+1\to n}$ is such that $\mathbb{P}_{U(n+1)} \circ \pi_{n+1\to n}^{-1} = \mathbb{P}_{U(n)}$ and the family of measures $\left(\mathbb{P}_{U(n)}\right)_n$ are projective and can hence be defined on the projective limit $\varprojlim_n (\mathcal{U}_n(\mathbb{C}), \pi_{n+1\to n})_n =: \mathfrak{U}(\mathbb{C})$. This space was defined by Bourgade, Najnudel and Nikeghbali in [14] as the space of *virtual rotations*. It generalizes the space of virtual permutations $\mathfrak{S}_{(\infty)}$ as well as the space of *virtual isometries* defined by Neretin in [78] [1].

Here again, the projections are not group morphisms and the space of virtual rotations is only a set. But a substitute of Haar measure in the action by conjugation of $\mathcal{U}_\infty(\mathbb{C}) := \varinjlim_n (\mathcal{U}_n(\mathbb{C}), \pi_{n+1\to n})_n$ on $\mathfrak{U}(\mathbb{C})$ is given by $\mathbb{P}_{\mathfrak{U}} := \varprojlim_n \mathbb{P}_{U(n)}$. As a caracterisation of $CUE(n) = (\mathcal{U}_n(\mathbb{C}), \mathbb{P}_{\mathcal{U}_n})$ is its invariance under the action of $\mathcal{U}_n(\mathbb{C})$ by conjugation, $(\mathfrak{U}(\mathbb{C}), \mathbb{P}_{\mathfrak{U}})$ can be thought of as a $CUE(\infty)$ (see [12]).

1.3 Questions of eigenvalues

1.3.1 The Weyl integration formula and its avatars

The invariance of a set under the action of a suitable group often allows to pass from a problem to a simpler one by considering the equivalence classes instead of the whole set. In the case of $\mathcal{U}_n(\mathbb{C})$ or $\mathfrak{S}_n$, the action of themselves by conjugation allows to consider the eigenvalues for a matrix of $\mathcal{U}_n(\mathbb{C})$ and the cycle structure for a permutation of $\mathfrak{S}_n$ (or another "coset structure" for another type of action).

The fact that every unitary matrix can be diagonalised with eigenvalues on the unit circle gives rise to a disintegration of the Haar measure of $\mathcal{U}_n(\mathbb{C})$ by means of the Haar measure of the group of diagonal matrices isomorphic to the torus $\mathbb{U}^n$ (with $\mathbb{U} := \mathbb{S}^1(\mathbb{C})$). Let $(e^{i\alpha_1}, \dots, e^{i\alpha_n})$ be the eigenangles of $U \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$, with $\alpha_k \in [0, 2\pi[$. The joint density of the eigenangles has been computed by H. Weyl (see [102]) and is given by the *Weyl denominator formula*

$$\mathbb{P}_{\mathcal{U}(n)} \left(\alpha_1 \in d\theta_1, \dots, \alpha_n \in d\theta_n\right) = \frac{1}{n!} \prod_{1\leqslant k<\ell\leqslant n} \left| e^{i\theta_k} - e^{i\theta_\ell} \right|^2 \prod_{k=1}^{n} \frac{d\theta_k}{2\pi} \tag{1.11}$$

This is a probability measure on $[0, 2\pi[^n$ that can be mapped into a probability measure on $\mathbb{U}^n$. Let us denote its Lebesgue-density by

$$f_n(\theta_1, \dots, \theta_n) = \frac{1}{(2\pi)^n \, n!} \prod_{1\leqslant k<\ell\leqslant n} \left| e^{i\theta_k} - e^{i\theta_\ell} \right|^2 = \frac{1}{(2\pi)^n \, n!} \left| \Delta\left(e^{i\theta_1}, \dots, e^{i\theta_n}\right) \right|^2$$

[1] This space is obtained by taking the image by the Cayley map of the projective limit of the Hermitian matrices for the projections that consist in deleting the last row and column, but due to the nature of the Cayley map, 1 is almost surely not an eigenvalue when endowed with the Haar measure, and it does not contain in particular the symmetric groups

where Δ is the Vandermonde determinant defined by

$$\Delta(z_1, \ldots, z_n) := \prod_{1 \leqslant k < \ell \leqslant n} (z_k - z_\ell)$$

This is the appearance of the Vandermonde determinant that explains all the phenomena that involve the eigenvalues of random unitary matrices. Indeed, all the information concerning the eigenvalues is contained in (1.11). Setting $Z_n := (2\pi)^n \, n!$, physicists are used to write

$$f_n(\theta_1, \ldots, \theta_k) = \frac{1}{Z_n} \exp\left(-2 \sum_{1 \leqslant k < \ell \leqslant n} -\log\left| e^{i\theta_k} - e^{i\theta_\ell} \right| \right) := \frac{1}{Z_n} \exp\left(-\beta \mathcal{H}(\theta_1, \ldots, \theta_n)\right) \Big|_{\beta=2}$$

with

$$\mathcal{H}(\theta_1, \ldots, \theta_n) := \sum_{1 \leqslant k < \ell \leqslant n} -\log\left| e^{i\theta_k} - e^{i\theta_\ell} \right|$$

This is the Hamiltonian or *interaction potential* of a *Coulomb gas* of n repelling electrical particles confined on the circle. Such a statistical mechanics interpretation can also be done for the Mallows measure since one can write (see e.g. [92])

$$\mathbb{M}_q^{(n)}(\sigma) = \frac{1}{Z_n(q)} \exp\left(-\ln(q^{-1}) \sum_{1 \leqslant k < \ell \leqslant n} \mathbb{1}_{\{\sigma_k - \sigma_\ell > 0\}} \right) =: \frac{1}{Z_n(q)} \exp\left(-\beta \widetilde{\mathcal{H}}(\sigma_1, \ldots, \sigma_n) \right) \Big|_{\beta=\ln(q^{-1})}$$

with $Z_n(q) := (1+q)(1+q+q^2)\ldots(1+q+\cdots+q^{n-1})$ and setting $H := \mathbb{1}_{\mathbb{R}_+^*}$

$$\widetilde{\mathcal{H}}(\sigma_1, \ldots, \sigma_n) = \sum_{1 \leqslant k < \ell \leqslant n} H(\sigma_k - \sigma_\ell)$$

In addition, one can map $[\![1, n]\!]$ on $\mathbb{U}$ and see a Mallows permutation as a particle system on $\mathbb{U}$ (but contrary to the eigenvalues of a random matrix which are "neatly spaced but slightly random" according to [30], the particles here are strictly spaced).

Remark 1.3.1. Another appearance of the Vandermonde determinant in an algebraic object is the dimension of an irreducible module of $\mathfrak{S}_n$: such a module V^λ is parametrized by a partition $\lambda \vdash n$ and its dimension is given by the *Young formula*

$$d_\lambda = \frac{\Delta(\lambda_1 - 1 + \frac{1}{2}, \lambda_2 - 2 + \frac{1}{2}, \ldots, \lambda_n - n + \frac{1}{2})}{\lambda_1! \, \lambda_2! \ldots \lambda_n!}$$

As in addition the following identity holds

$$\sum_{\lambda \vdash n} \frac{d_\lambda^2}{n!} = 1$$

Denote by $\mathbb{Y}_n := \{\lambda \vdash n\}$ the set of partitions of size n that we can identify with the set of Young diagrams. Using this last identity, one can define a probability measure $\mathcal{P}_n$ on $\mathbb{Y}_n$ by setting $\mathcal{P}_n(\lambda) := d_\lambda^2 / n!$ (Plancherel measure) and for $\lambda \sim \mathcal{P}_n$, if we define λ' by $\lambda'_k := \lambda_k - k + 1/2$, then λ' can be tought of as a particle system on $\mathbb{Z} + 1/2$.

Due to the appearance of the power 2 of the Vandermonde determinant, the parts $(\lambda_1, \lambda_2, \dots)$ of such a random partition $\lambda \sim \mathcal{P}_n$ share some similarities with the eigenvalues of $U \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$, especially in the bulk where the Poissonized version of the Plancherel measure forms a determinantal point process with discrete sin kernel (see e.g. [80, 56]). But since the parts of a partitions are on a line, their good analogue are the eigenvalues of $H \sim GUE(n)$. One of the most striking resemblance between the parts of a random Plancherel partition and the eigenvalues of a matrix of the $GUE(n)$ is the fact that they share the same fluctuations of their edge statistics (i.e. the largest part and the largest eigenvalue, fluctuate according to the Tracy-Widom distribution). This result was extended to all higher order statistics (see [56]).

1.3.2 Joint intensities and determinantal processes

How to describe a point process ? Sometimes, like in the case of the Poisson point process, the description is straightforward due to a lot of independence.

Another example can be given by considering the case of the Ewens measure ; denote by $\mathcal{C}$ the *cycle structure* $\mathcal{C}(\sigma) := (c_k(\sigma))_{1 \leqslant k \leqslant n}$ where $c_k(\sigma)$ denotes the number of k-cycles of σ. Under the Ewens measure, the cycle structure $\mathcal{C}$ is a sequence of random variables that can be thought of as system of particles on $[\![1, n]\!]$. The random variables X_n obtained by a reordering of $\mathcal{C}/\sum_{k\geqslant 1} c_k$ is a random point process in $(0, 1]$ and is shown to converge to a locally finite point process $Y = \sum_k \delta_{Y_k}$ on $(0, 1]$ by *Kingman's theorem* (see [66]). This limiting process is called *Poisson-Dirichlet process* and can be described by means of an explicit operation on a Poisson point process on the real line with a given intensity (see [2]). But another description can be given by the following function

$$\mathbb{P}\left(Y\left(dy_1\right) \neq 0, Y\left(dy_2\right) \neq 0, \dots, Y\left(dy_k\right) \neq 0\right) = \rho_k(y_1, \dots, y_n) dy_1 \dots dy_2$$

which is given explicitely by (see [101] or [2] pp. 82)

$$\rho_k(y_1, \dots, y_k) = \frac{\theta^k}{y_1 \dots y_k} \left(1 - \sum_{\ell=1}^{k} y_\ell\right)_+^{\theta-1} \mathbb{1}_{\{1 \geqslant y_1 \geqslant y_2 \geqslant \dots \geqslant y_k \geqslant 0\}}$$

Such a function ρ_k is said to be the k-joint intensity of the point process Y (or k-correlation of Y). More generally, we define (for all the definitions of this paragraph, see e.g. [1, 9])

Definition 1.3.2 (Joint intensities)**.** Let Y be a simple point process on a locally compact Polish space Λ, that is, a random integer-valued positive Radon measure that almost surely assigns measure 0 or 1 to singletons. Let μ be a measure on Λ.

The joint intensities of Y with respect to μ are the functions $\rho_k : \Lambda^k \to \mathbb{R}_+$ given, if they exist, by : for all mutually disjoint subsets $D_1, \dots, D_k \subset \Lambda$

$$\begin{aligned} &\mathbb{E}\left(\prod_{\ell=1}^{k} Y(D_\ell)\right) = \int_{\prod_{\ell=1}^{k} D_\ell} \rho_k(y_1, \dots, y_k) d\mu^{\otimes k}(y_1, \dots, y_k) \\ &\rho_k(y_1, \dots, y_k) = 0 \ \text{ if } \exists\, i \neq j \text{ s.t. } x_i = x_j \end{aligned}$$

Note that we have not defined ρ_k by $\mathbb{E}\left(Y^{\otimes k}(A)\right)$ for a certain $A \subset \Lambda^k$ due to the problem of definition on the diagonal. Indeed, ρ_k would be the intensity measure of $Y^{\wedge k}$, the point process of the set of ordered k -tuples of distinct points of Y. From this point of view, a good intuition is given by the formula

$$\rho_k(y_1, \dots, y_k) = \lim_{\varepsilon \to 0} \frac{\mathbb{P}\left(Y(B_\varepsilon(y_j)) \neq 0,\ \forall j \in [\![1,k]\!]\right)}{\mu(B_\varepsilon)^k} = \frac{\mathbb{P}\left(Y(dy_j) \neq 0,\ \forall j \in [\![1,k]\!]\right)}{dy_1 \dots dy_k}$$

Note also that if $Y = \sum_{k=1}^n \delta_{Y_k}$ where $(Y_1, \dots, Y_n)$ are exchangeable real valued random variables with joint density $p(x_1, \dots, x_n)$ with respect to the Lebesgue measure on $\mathbb{R}^n$, then, the k-joint intensities are proportional to the k-joint marginal distribution of $(Y_1, \dots, Y_n)$ in the sense that

$$\rho_k(y_1, \dots, y_k) = \frac{n!}{(n-k)!} \int_{\mathbb{R}^{n-k}} p(x_1, \dots, x_n) dx_{k+1} \dots dx_n \tag{1.12}$$

One can ask if these joint intensities determine the distribution of a point process. The answer is yes under some restrictions (see e.g. [9] p. 9).

Example 1.3.3. When $k = 1$, the first (joint) intensity can be thought of as the density of the point process Y since

$$\mathbb{E}\left(Y(A)\right) = \int_A \rho_1(y) d\mu(y)$$

Example 1.3.4. When Λ is countable, the joint intensities are given by

$$\rho_k(x_1, \dots, x_k) = \mathbb{P}\left(Y(\{x_1, \dots, x_k\}) \neq 0\right)$$

In the case of the eigenvalues of $U \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$, the key class of point processes is the following :

Definition 1.3.5 (Determinantal point process)**.** Let Y be a simple point process on a locally compact Polish space Λ and let μ be a measure on Λ. Let $K : \Lambda^2 \to \mathbb{C}$ be a measurable function.

Y is said to be a determinantal point process of kernel K if its k-joint intensities ρ_k (with respect to μ) can be written for all $(x_1, \dots, x_k) \in \Lambda^k$ as

$$\rho_k(x_1, \dots, x_k) = \det\left(K(x_i, x_j)\right)_{1 \leqslant i,j \leqslant k}$$

The advantages of such processes are due to several formulas available for algebraic manipulations. For example,

- Suppose that the kernel K satisfies the two following properties

  1. Autoreproduction : $\int_\Lambda K(x,y)K(y,z)d\mu(y) = K(x,z)$,
  2. Trace : $\int_\Lambda K(x,x)d\mu(x) = n$

this last property being proved to be equivalent to have n particles almost surely. Then, we have the formula

$$\int_\Lambda \det\left(K(x_i,x_j)\right)_{1\leqslant i,j\leqslant n} d\mu(x_n) = (n-1)\det\left(K(x_i,x_j)\right)_{1\leqslant i,j\leqslant n-1}$$

and by induction, all the k-joint intensities follow.

- If $Y = \sum_{k\geqslant 1}\delta_{Y_k}$ is a point process, the following identity holds

$$\begin{aligned}\mathbb{E}\left(\prod_{k=1}^n (1+f(Y_k))\right) &= \mathbb{E}\left(\sum_{k=0}^n e_k\left(f(Y_1),\dots,f(Y_n)\right)\right)\\ &= \sum_{k=0}^n \frac{1}{k!}\int_{\Lambda^k}\rho_k(x_1,\dots,x_k)f(x_1)\dots f(x_k)d\mu^{\otimes k}(x_1,\dots,x_k)\end{aligned}$$

where $e_k(z_1,\dots,z_n)$ is the k-th elementary symmetric polynomial defined by

$$e_k(z_1,\dots,z_n) := \sum_{1\leqslant i_1<i_2<\dots<i_k\leqslant n} z_{i_1}z_{i_2}\dots z_{i_k}$$

In the case of a determinantal point process of kernel K, this identity becomes

$$\mathbb{E}\left(\prod_{k=1}^n (1+f(Y_k))\right) = \sum_{k=0}^n \frac{1}{k!}\int_{\Lambda^k}\det\left(K(x_i,x_j)\right)_{i,j\leqslant k} f(x_1)\dots f(x_k)d\mu^{\otimes k}(x_1,\dots,x_k)$$

and we recognise the beginning of a Fredholm development, namely, if the operator $\boldsymbol{K}$ of kernel K acts on $L^2(\Lambda, f\bullet\mu)$ by $\boldsymbol{K}g(x) := \int_\Lambda K(x,y)g(y)d(f\bullet\mu)(y)$ for all $g\in L^2(\Lambda,f\bullet\mu)$, we have

$$\det(I+\boldsymbol{K}) = \sum_{k\geqslant 0}\frac{1}{k!}\int_{\Lambda^k}\det\left(K(x_i,x_j)\right)_{i,j\leqslant k} f(x_1)\dots f(x_k)d\mu^{\otimes k}(x_1,\dots,x_k)$$

In particular, taking $f = -\mathbb{1}_A$, we get the following identity

$$\mathbb{P}\left(\forall\, k\leqslant n,\, Y_k\notin A\right) = \sum_{k=0}^n \frac{(-1)^k}{k!}\int_{A^k}\det\left(K(x_i,x_j)\right)_{i,j\leqslant k} d\mu^{\otimes k}(x_1,\dots,x_k)$$

that allows to compute extreme values probabilities and that led to the Tracy-Widom and Gaudin-Metha distributions, whose probability write respectively as the Airy and sin kernel acting on a suitable $L^2([x,+\infty[)$ space.

Back to the $CUE(n)$, it is easily checked with the joint density (1.11) and the formula (1.12) that for $U\sim\mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$, the point process of the eigenvalues of U is determinantal. More precisely, we have the

Theorem 1.3.6 (Dyson, [34]). *Let $U \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$ with spectrum $\left\{e^{i\theta_{k,n}}\right\}_{1\leqslant k\leqslant n}$. Then, the point process $\Theta_n := \sum_{k=1}^n \delta_{\exp(i\theta_{k,n})}$ is determinantal of kernel*

$$K_n\left(e^{i\theta}, e^{i\phi}\right) = \sum_{k=0}^{n-1} e^{ik(\theta-\phi)} = e^{i(n-1)(\theta-\phi)/2}\frac{\sin\left(\frac{n}{2}(\theta-\phi)\right)}{\sin\left(\frac{\theta-\phi}{2}\right)}$$

with respect to the (normalised) Lebesgue measure on $[0, 2\pi)$ (with total mass 1).

Note that this kernel is the kernel of the projection operator $P_n : L^2(\mathbb{U}) \to \mathrm{span}(e_0, \dots, e_{n-1})$ with $e_k(\theta) := e^{ik\theta}$, that is, for all $f \in L^2(\mathbb{U})$, $P_n f\left(e^{i\theta}\right) = \int_0^{2\pi} K_n\left(e^{i\theta}, e^{i\phi}\right) f\left(e^{i\phi}\right) \frac{d\phi}{2\pi}$.

Note also that a kernel is only defined up to certain multiplications, since for all $f : \mathbb{R} \to \mathbb{C}^\times$

$$\det\left(K(x_i, x_j)\right)_{i,j\leqslant n} = \det\left(\frac{f(x_i)}{f(x_j)}K(x_i, x_j)\right)_{i,j\leqslant n}$$

In particular, one can consider that $(\Theta_n)_n$ is determinantal of kernel

$$\widetilde{K}_n\left(e^{i\theta}, e^{i\phi}\right) = \frac{\sin\left(\frac{n}{2}(\theta-\phi)\right)}{\sin\left(\frac{\theta-\phi}{2}\right)}$$

It is moreover clear that $\widetilde{K}_n(x/n, y/n) \to \sin(\pi(x-y))/(\pi(x-y))$ when $n \to +\infty$ (for $x = y$, the function $x \to \sin(x)/x$ is extended in 0 by the value 1). This limit can be interpreted by means of the convergence of the process of the rescaled eigenangles $\sum_{k=1}^n \delta_{n\theta_{k,n}}$ to a general determinantal process on the circle with kernel

$$K(x, y) = \frac{\sin\left(\pi(x-y)\right)}{\pi(x-y)}$$

1.3.3 Cycles of a random permutation and trace of powers of a random matrix

On the symmetric group $\mathfrak{S}_n$, fruitful questions about what look like typical elements of the group involve questions on functionals of random permutations selected from the uniform measure. One can ask for instance questions such as "how many cycles in a permutation", "what is the order of a permutation" (minimal power of the permutation to get the identity), etc. Answers to these questions involve the number of cycles $C(\sigma_n)$ of a random permutation $\sigma_n \sim \mathbb{P}_1^{(n)}$ or its order $\mathrm{ord}(\sigma_n)$ and can be stated by means of the following convergence in distribution

$$\begin{aligned} \frac{C(\sigma_n) - \log(n)}{\sqrt{\log(n)}} &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}(0,1) \quad \text{(Goncharov)} \\ \frac{\mathrm{ord}(\sigma_n) - (\log n)^2/2}{(\log(n)/3)^{3/2}} &\xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}(0,1) \quad \text{(Erdös-Turan)} \end{aligned}$$

Each of these results can be proven by a suitable analysis of the cycle structure of the permutation, using the Feller coupling (see [2]).

Cycle statistics of a permutation can be thought of as equivalents of eigenvalues statistics for a matrix, since the eigenvalues of a random permutation are coded by the cycle structure. One can dress a parallel between those two worlds to see the resemblance.

As listed in [30], fundamental facts about the eigenvalues of $U_n \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$ are the following :

- **Neat spacing :** The following convergence in distribution with no renormalisation

$$\mathrm{tr}(U_n) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}_{\mathbb{C}}(0,1)$$

implies that the eigenvalues on the unit circle are closed to n equally spaced numbers (contrary to n uniform independent numbers that are of order $\sqrt{n}$ by the central limit theorem). A closer look at this phenomenon can be made by computing the number of eigenvalues in a given arc. This was done by Wieand [103] that proves the following

$$\frac{1}{\sqrt{\log(n)}}\left(\sum_{k=1}^{n} \mathbb{1}_{\left\{a\leqslant\theta_{k,n}\leqslant b\right\}} - n(b-a)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}(0,1) \tag{1.13}$$

Hence, small fluctuations of order $\sqrt{\log(n)}$ can complete the picture of the spacing. Note that this result also holds for the eigenangles of a uniform random permutation (see [104]) and Ewens permutations (see [16, 50]).

- **Traces are almost Gaussian :** The classical Berry-Esséen theorem in proability shows that for sum of i.i.d. random variables, the difference made by approximating the renormalised sum by its Gaussian counterpart is of order $1/\sqrt{n}$. This phenomenon is slightly improved on the unit circle (order $1/n$) but is far from achieving the phenomenon obtained for the eigenvalues of a random unitary matrix : setting $Z \sim \mathscr{N}(0,1)$, there exist universal constants $c, \sigma > 0$ such that

$$\sup_{B \ \text{Borel}} \left|\mathbb{P}\left(\mathrm{tr}(U_n)\in B\right) - \mathbb{P}\left(Z \in B\right)\right| \leqslant \frac{c}{n^{\sigma n}} \tag{1.14}$$

This result due to Johansson (see [55]) has its counterpart in the world of uniform permutations, a result that goes back to Monmort in 1708, the trace of P_σ being the number of its fixed points FP :

$$\mathbb{P}_1^{(n)}\left(FP = k\right) = \frac{1}{e\,k!} + O\left(\frac{2^n}{(n+1)!}\right)$$

With $X \sim \mathscr{P}(1)$, this result writes

$$\sup_{B\subset[\![1,n]\!]} \left|\mathbb{P}\left(FP\in B\right) - \mathbb{P}\left(X \in B\right)\right| \underset{n\to\infty}{\sim} \frac{2^n}{(n+1)!}$$

- **Traces of successive powers are still Gaussian, but less close to it :** A celebrated result of Diaconis and Shashahani (see [33]) states that

$$\mathrm{tr}(U_n^k) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}_{\mathbb{C}}(0,k) \tag{1.15}$$

And in addition, for all $\ell \in \mathbb{N}$, we have independence between the traces at the limit :

$$\left(\frac{\mathrm{tr}(U_n^k)}{\sqrt{k}}\right)_{k\leqslant \ell} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}_{\mathbb{C}}(0,I_\ell)$$

Nevertheless, the speed of convergence becomes less and less fast, and a phase transition occurs for $k=n$, in virtue of the following theorem due to Rains (see [82]) : for $k \geqslant n$, the eigenvalues of $\mathrm{tr}(U_n^k)$ are exactly distributed as n i.i.d. random variables of uniform law on the unit circle. According to the classical Berry-Esséen theorem in this case, the total variation distance between the sum of these eigenvalues and the Gaussian distribution is then of order $1/n$.

This theorem is in fact general for any polynomial distribution on the unit circle (see [9] p. 74). One can easily check that if $(X_k)_k$ is a sequence of i.i.d. uniform random variables on $[0,1]$, we have the following convergence without any renormalisation

$$\sum_{k=1}^{n} e^{2i\pi(X_k+k)/n} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}_{\mathbb{C}}(0,1) \tag{1.16}$$

Hence, from the point of view of the Central limit theorem, the eigenvalues of a random unitary matrix mimic the behaviour of n-th roots of n i.i.d. uniform random numbers on the unit circle, the n-th root being taken in such a way that the neat spacing is preserved (i.e. $[e^{2i\pi X_k}]^{1/n} := e^{2i\pi(X_k+k)/n}$ but another choice is possible). It would be interesting to go beyond this result, and, in the flavour of mod-$*$ convergence (see definition 3.2.1), compare the distributions of the traces and of these n-th roots of i.i.d. uniform numbers on the circle at the second order, or even to recreate a random n-th rooting that gives the law of the eigenvalues starting from i.i.d's.

Of course, the same result occurs for traces of powers of a random uniform permutation (see e.g. [2]).

Remark 1.3.7. The Diaconis-Shashahani theorem on the convergence of the traces of powers of $U_n \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$ is equivalent to the strong Szegö theorem on the asymptotic behaviour of Toeplitz determinants (see [54] or [30]).

1.4 The characteristic polynomial of a random unitary matrix

1.4.1 Correlations of the number of eigenangles in an arc

As pointed in (1.13) The number of eigenangles in a prescribed arc is approximately Gaussian.

Wieand [103] also calculated the correlations of the Gaussian limiting process $(Z_{a,b})_{a<b}$ of $Z_{a,b}^{(n)} := \left(\sum_{k=1}^{n} \mathbb{1}_{\{a \leqslant \theta_{k,n} \leqslant b\}} - n(b-a)/2\pi\right) / \sqrt{\log(n)/\pi^2}$ and found, for $a<b$ and $c<d$

$$\mathbb{E}\left(Y_{a,b}Y_{c,d}\right) = \begin{cases} 0 & \text{if} \quad \partial]a,b[\cap\partial]c,d[=\varnothing \\ -\frac{1}{2} & \text{if} \quad b=c \\ \frac{1}{2} & \text{if} \quad a=c \end{cases} \tag{1.17}$$

These correlations say that if $[a,b]$ contains $[c,d]$ or if $[a,b]$ and $[c,d]$ are disjoint, the random variables are independent, while if $[a,b]$ and $[c,d]$ share an endpoint, the variables have correlation $\pm 1/2$.

A way to obtain these correlations is the following : define $X(f) := \int_0^{2\pi} f dX$ for $f = \mathbb{1}_{[a,b]}$ and X a $1/2$-white noise, that is, the weak $1/2$-fractional differential of the Brownian motion (taken in the sense of distributions). This amounts to assign a collection of i.i.d. Gaussian random variables on every point of the unit circle, i.e. $X(\mathbb{1}_{[a,b]}) = Z_b - Z_a$ where the $(Z_a)_{a\in[0,2\pi[}$ are i.i.d. Gaussian random variables with variance $1/2$. One can check that the covariance of this process is the one given in (1.17). Hence, the linear statistics $Z_{a,b} = \int_0^{2\pi} \mathbb{1}_{[a,b]} d\mu_n$ where $\mu_n := \sum_{k=1}^n \delta_{\theta_{k,n}}$ behave like an integrated white noise $X(\mathbb{1}_{[a,b]}) := \int_0^{2\pi} \mathbb{1}_{[a,b]} dX$, asking the question of the nature of this white noise.

The answer was given by Hughes and al. in [49] : it is in the characteristic polynomial that lies the mystery. One can indeed write, for $|z|<1$

$$\Phi_U(z) := \det\left(I_n - zU\right) = \exp\left(\operatorname{tr}\log\left(I_n - zU\right)\right) = \exp\left(-\sum_{k\geqslant 1}\frac{\operatorname{tr}(U^k)}{k} z^k\right)$$

According to the Diaconis-Shashahani theorem (1.15), when $n\to\infty$, $\operatorname{tr}(U^k)/\sqrt{k}$ are approximately i.i.d. Gaussians, the log-characteristic polynomial has the same asymptotic distribution as

$$\sum_{k\geqslant 1}\frac{Z_k}{\sqrt{k}} z^k$$

where $(Z_k)_k$ is an i.i.d. sequence of Gaussian random variables.

For $|z|<1$, the last series converges uniformly and one can write explicitely the covariance. But on the circle $|z|=1$, the white noise is not well defined (the variance is equal to $+\infty$ while the extra-diagonal covariance is finite) and one has to renormalise to find the Gaussian limit

$$\frac{\log\Phi_U\left(e^{i\theta}\right)}{\sqrt{\log(n)/\pi^2}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}_{\mathbb{C}}(0,1)$$

Integrating on θ on a suitable arc (a,b), one can find the result of Wieand (for the details, see [103]). The additional renormalisation that consists in substracting $n(b-a)/(2\pi)$ is a general fact of linear statistics with discontinuous test function.

1.4.2 The link with number theory

The Riemann Zeta function is defined for $\mathfrak{Re}(s) > 1$ by its Dirichlet series or its Eulerian product indexed by the set $\mathcal{P}$ of prime numbers

$$\zeta(s) := \sum_{n \geqslant 0} \frac{1}{n^s} = \prod_{p \in \mathcal{P}} \frac{1}{1 - p^{-s}} \tag{1.18}$$

This function can be extended meromorphically to the whole complex plane thanks to the following functional equation :

$$\zeta(s) = \pi^{s-1/2} \frac{\Gamma\left(\frac{1-s}{2}\right)}{\Gamma\left(\frac{s}{2}\right)} \zeta(1-s)$$

Introducing the function $\xi : s \mapsto \pi^{-s/2}\Gamma(s/2)\zeta(s)$, one can rewrite this equation into

$$\xi(s) = \xi(1-s)$$

One can notice that ξ is the Mellin transform of a random variable (see e.g. [18]).

The key point about ζ lies in its Eulerian product : it encodes the whole structure of the prime set $\mathcal{P}$. Thus, a particular information about ζ gives a particular information about the very structure of the prime sequence. For example, it is known since Euclide that the set of primes is infinite, and one can ask about the speed of apparition of the primes amongst the integer. This result first conjectured by Gauss and Legendre is the celebrated prime number theorem due to Hadamard and de la Valée-Poussin (1896) that makes a considerable use of complex analysis computations on $\zeta(s)$ and that can be stated as

$$\pi_n := \sum_{p \in \mathcal{P}} \mathbb{1}_{\{p \leqslant n\}} \underset{n \to \infty}{\sim} \frac{n}{\log(n)}$$

Riemann was the first to point that except the trivial zeroes at $s = -2, -4, \ldots$, all the zeros lie in the critical strip $\{0 < \mathfrak{Re} < 1\}$. The conjectural fact that all the zeros lie on the critical line $\{\mathfrak{Re} = 1/2\}$ is the Riemann hypothesis, that would imply, after a carefull analysis, a refinement of the latest speed of convergence into

$$\pi_n = \int_2^n \frac{dt}{\log t} + O\left(\frac{1}{n^{1/2+\varepsilon}}\right), \quad \forall \varepsilon > 0$$

A first statistical look at the zeros of ζ was first done by Selberg in his celebrated Central limit theorem (see e.g. [97]).

Theorem 1.4.1 (Selberg's Central limit theorem)**.** *Let $U \sim \mathscr{U}([0,1])$ (or $[1,2]$) and $T > 0$. Then, the following convergence in distribution holds*

$$\frac{\log\left|\zeta\left(\frac{1}{2} + iTU\right)\right|}{\sqrt{\frac{1}{2}\log\log T}} \xrightarrow[T \to +\infty]{\mathcal{L}} \mathscr{N}(0,1)$$

Thus, $\left|\zeta\left(\frac{1}{2}+iT\,U\right)\right|$ is approximately a log-normal distribution of variance $\frac{1}{2}\log\log T$, which can be restated into a result about the non probabilistic values of ζ on the critical line, of order $O(1)$ plus some small error term. Note the contrast with the behaviour outside the critical axis, where a convergence in law holds without any renormalisation according to the *Bohr-Jessen theorem* (see e.g. [67] and references cited) : for all $\varepsilon > 0$

$$\left|\zeta\left(\frac{1}{2}+\varepsilon+iT\,U\right)\right| \xrightarrow[T\to+\infty]{\mathcal{L}} \prod_{p\in\mathcal{P}}\left|1+\frac{e^{2i\pi U_p}}{p^{\frac{1}{2}+\varepsilon}}\right|^{-1}$$

where $(U_p)_p$ is a sequence of i.i.d. random variables uniformly distributed on $[0,1]$.

A second statistical result of interest concerning the distribution of the zeroes of ζ lies in the following conditional result due to Montgommery concerning the pair correlations or 2-joint intensities of its zeroes

Conjecture 1.4.2 (Montgommery, [75]). Suppose the Riemann Hypothesis, denote the non trivial zeroes of ζ on $\{\mathfrak{Im}\geqslant 0\}$ by $\left(\frac{1}{2}+\mathrm{i}\gamma_k\right)_{k\geqslant 1}$ with $\gamma_k\geqslant 0$ and set $\widehat{\gamma}_k := \gamma_k\log\left(\frac{\gamma_k}{2\pi}\right)$ (so that $|\widehat{\gamma}_{k+1}-\widehat{\gamma}_k|\sim 1$). Then

$$\frac{1}{N}\sum_{1\leqslant k\neq \ell\leqslant N}\phi\left(\widehat{\gamma}_k-\widehat{\gamma}_\ell\right) \xrightarrow[N\to+\infty]{} \int_{\mathbb{R}}\phi(x)\left(1-\left(\frac{\sin(\pi x)}{\pi x}\right)^2\right)dx$$

holds for $\mathrm{supp}(\mathcal{F}\phi)\subset[a,b]$ for all $a,b\in\mathbb{R}$ (with $\mathcal{F}$ the Fourier transform).

Montgommery proved in fact the following

Theorem 1.4.3 (Montgommery, [75]). *The last limit holds for $\phi\in\mathcal{C}^\infty$ with* $\mathrm{supp}(\mathcal{F}\phi)\subset[-1,1]$.

A first point to notice is that the conjecture about the extension of the support is extremely hard (for $[1-\varepsilon,1+\varepsilon]$, it involves a quantitative version of the twin primes conjecture.

The key point to remark is the following : the 2-joint intensities of the zeros of ζ are the same as the (renormalised) 2-joint intensities 1.3.6 of the eigenvalues of a CUE matrix computed by Dyson[2]. This statistical correspondance must be precised to avoid a simple coincidence.

If the zeroes of the Riemann Zeta function behave like the zeroes of the characteristic polynomial of a CUE matrix of large size, what can be said about the Zeta function on the critical line and the characteristic polynomial on the unit circle ? The celebrated Keating-Snaith philosophy asserts that $Z_N(\theta)$ for large N is a toy model of $\zeta(1/2+it)$. The conjectural fact that $\zeta(1/2+it)$ should be the characteristic polynomial of a certain self-adjoint operator (hence having all its eigenvalues real) is the Hilbert-Polya conjecture, that has proven to be

[2] According to the official anecdote, this connection was made during a tea break where Montgommery met the physicist Freeman Dyson. As Montgomery relates it, " I suppose that by now somebody else would have seen the connection... it's nearly thirty years ago. But it certainly was, from the standpoint of publication, instantaneous. I had the mathematics and as soon as I had it, it was just a matter of months before the connection was pointed out". (from : K. Sabbagh, *Dr. Riemann's Zeros*, Atlantic, 2002, pp. 134-136)

true in a previous analogous case : the Zeta functions over a function field, for instance $\mathbb{F}_q[X]$ with $q = p^\nu$ for $p \in \mathcal{P}$. These functions can be defined as

$$\zeta_{\mathbb{F}_q[X]}(s) := \prod_{\pi \in \mathcal{P}(\mathbb{F}_q[X])} \frac{1}{1 - |\pi|_q^{-s}}$$

where $\mathcal{P}(\mathbb{F}_q[X])$ denotes the set of prime polynomials (i.e. with no divisors) and $|\pi|_q := q^{\deg(\pi)}$. This Zeta function is a direct generalisation of the classical one, that can be seen as $\zeta_{\mathbb{N}}$. The same kind of generalisation can be made for structures having a divisibility property : permutations, graphs, posets, $\mathbb{Z}[i]$, polynomials, closed geodesics on manifolds, etc.

The Weil conjectures, finally proven in full generality by Grothendieck, asked the question of their rationality and their representation as the characteristic polynomial of a certain operator. The proof given by Grothendieck consists precisely in constructing the operator : this is the Frobenius map $x \mapsto x^p$ acting on a certain cohomology space of the variety (the so-called *étale cohomology* whose construction was one achievement of modern algebraic geometry). The fact that Zeta functions over function fields have all their roots on the critical line $\{\mathfrak{Re} = 1/2\}$ was proven by Deligne, ending the Weil program. In 1999, Katz and Sarnak proved the Montgomery conjectures on function fields ([60]) giving additional evidence about the conjecture for ζ.

Back to the classical Riemann zeta function on $\mathbb{N}$; a more refined link between $Z_N(\theta)$ and $\zeta(1/2+it)$ consists in looking at what happens at the second order of renormalisation. Selberg proved in fact a more general statement : the convergence in his theorem occurs in moments, i.e. the integer moments of $\log|\zeta(1/2 + iT\,U)|$ converge to those of a gaussian distribution after renormalisation :

$$\mathbb{E}\left(\left|\log\zeta\left(\frac{1}{2} + iT\,U\right)\right|^{2k}\right) = \left(\frac{1}{2}\log\log T\right)^{k} \mathbb{E}\left(G^{2k}\right)(1+o(1)) \quad \text{with } G \sim \mathscr{N}(0,1) \tag{1.19}$$

In particular, this theorem implies Selberg's CLT

$$\frac{\log\left|\zeta\left(\frac{1}{2} + iT\,U\right)\right|}{\sqrt{\frac{1}{2}\log\log T}} \xrightarrow[T\to+\infty]{\mathcal{L}} \mathscr{N}(0,1)$$

A more refined development is conjecturally the following (see [62]) : for all $\lambda \in \{\mathfrak{Re} > -1\}$

$$\mathbb{E}\left(\left|\zeta\left(\frac{1}{2} + iT\,U\right)\right|^{2\lambda}\right) = (\log T)^{\lambda^2} M(\lambda) A(\lambda)\,(1+o(1)) \tag{1.20}$$

where $A(\lambda)$ is the *arithmetic factor* given by

$$A(\lambda) := \prod_{p\in\mathcal{P}} \left[\left(1 - \frac{1}{p}\right)^{\lambda^2} \left(\sum_{k\geqslant 0} \left(\frac{\lambda(\lambda+1)\cdots(\lambda+k-1)}{k!\,p^{k/2}}\right)^2\right)\right]$$

and where $M(\lambda)$ is the *random matrix factor* given by

$$M(\lambda) = \frac{(\mathcal{G}(1+\lambda))^2}{\mathcal{G}(1+2\lambda)}$$

by means of the Barnes G-function defined for all $z \in \mathbb{C}$ by

$$\mathcal{G}(z+1) := (2\pi)^{z/2} e^{-[(1+\gamma)z^2+z]/2} \prod_{n \geqslant 1} \left(1+\frac{z}{n}\right)^n e^{-z+(z^2/2n)} \tag{1.21}$$

γ being here the Euler constant. Note that the G-function also satisfies the functional equation $\mathcal{G}(z+1) = \Gamma(z)\mathcal{G}(z)$.

The development (1.20) is only proven for $\lambda = 1$ (Hardy-Littlewood) and $\lambda = 2$ (Ingham). The general form given in (1.20) is the celebrated *moments conjecture*. One can rewite this equality in terms of Laplace transform instead of Mellin transform to get, with $G \sim \mathscr{N}(0,1)$ and for all $\lambda \in \{\mathfrak{Re} > -1\}$

$$\frac{\mathbb{E}\left(e^{\lambda \log\left|\zeta\left(\frac{1}{2}+iTU\right)\right|^2}\right)}{\mathbb{E}\left(e^{\lambda G\sqrt{2\log\log T}}\right)} = M(\lambda)A(\lambda)\left(1+o(1)\right) \tag{1.22}$$

With this form, one can see directly that (1.22) implies (1.19) since it amounts to change the normalisation, if nevertheless the convergence holds locally uniformly in λ, which is hidden in the $o(1)$, and if the limiting function $\lambda \mapsto A(\lambda)M(\lambda)$ is continuous, which can be proven to be the case. One says that a sequence of random variables converges *in the mod-Gaussian sense* precisely if this type of convergence holds locally uniformly (see definition 3.2.1).

The important fact in the Keating-Snaith philosophy is that the appearance of $M(\lambda)$ is due to the following equality

$$\mathbb{E}\left(\left|\det\left(I_n - U\right)\right|^{2\lambda}\right) = n^{\lambda^2} M(\lambda)\left(1+o(1)\right) \tag{1.23}$$

i.e. from the characteristic polynomial of a $CUE(n)$ matrix taken in one point of the unit circle (namely 1). An explanation of this replacement will be given in the next chapters. For the moment, let us consider that this characteristic polynomial as a function on the unit circle is a toy-model for the Riemann Zeta function on the critical line, which motivates the study of this random variable in more details.

1.4.3 The characteristic polynomial at one point of the unit circle

Define the characteristic polynomial on the unit circle by

$$Z_U(\theta) := \Phi_U\left(e^{-i\theta}\right)$$

As $Z_U(\theta) \stackrel{\mathcal{L}}{=} Z_{V^*U_NV}(\theta)$ for all $V \in \mathcal{U}_N(\mathbb{C})$, we have

$$Z_U(\theta) \stackrel{\mathcal{L}}{=} Z_U(0)$$

Defining the logarithm with its principal branch (a cut on $\mathbb{R}_-$), the bidimensional Mellin transform of $Z_U(0)$ is defined, for all $s,t \in \mathbb{C}$ with $\mathfrak{Re}(s \pm t) > -1$ by

$$\begin{aligned}
&\mathbb{E}_{\mathcal{U}(n)}\left(e^{s\,\mathfrak{Re}\log Z_X(0)+t\,\mathfrak{Im}\log Z_X(0)}\right)\\
&\quad = \frac{1}{n!}\int_{[0,1]^n}\prod_{k=1}^n \left|1-e^{2i\pi\theta_k}\right|^s e^{t\arg\left(1+e^{2i\pi\theta_k}\right)}\left|\Delta\left(e^{2i\pi\theta_1},\dots,e^{2i\pi\theta_n}\right)\right|^2 d\boldsymbol{\theta}
\end{aligned}$$

This last transform is given by

$$\mathbb{E}_{\mathcal{U}(n)}\left(e^{s\,\mathfrak{Re}\log Z_X(0)+t\,\mathfrak{Im}\log Z_X(0)}\right)=\prod_{k=1}^{n}\frac{\Gamma(k)\Gamma(k+s)}{\Gamma\left(k+\frac{s+it}{2}\right)\Gamma\left(k+\frac{s-it}{2}\right)} \tag{1.24}$$

Here, we have used the classical probabilistic convention to write X for the canonical evaluation $X(\omega)=\omega$ for $\omega\in\Omega:=\mathcal{U}_n(\mathbb{C})$.

One can rewrite equality (2.6) using the Barnes G-function (1.21) into

$$\mathbb{E}_{\mathcal{U}(n)}\left(e^{s\,\mathfrak{Re}\log Z_X(0)+t\,\mathfrak{Im}\log Z_X(0)}\right)=\frac{\mathcal{G}\left(1+\frac{s+it}{2}\right)\mathcal{G}\left(1+\frac{s-it}{2}\right)\mathcal{G}\left(1+n\right)\mathcal{G}\left(1+n+s\right)}{\mathcal{G}\left(1+n+\frac{s+it}{2}\right)\mathcal{G}\left(1+n+\frac{s-it}{2}\right)\mathcal{G}\left(1+s\right)} \tag{1.25}$$

Equality (2.6) was first computed in [62] using the following formula due to Selberg (see [87]) and a change of variables for the case $\gamma=1$.

Theorem 1.4.4 (Selberg integral)**.** *The following equality holds*

$$\int_{\mathbb{R}^n}\prod_{k=1}^{n}(a+ix_k)^{-\alpha}(b-ix_k)^{-\beta}\left|\Delta(\boldsymbol{x})\right|^{2\gamma}d\boldsymbol{x}\quad= \tag{1.26}$$

$$\frac{(2\pi)^{2n}}{(a+b)^{(\alpha+\beta)n-\gamma n(n-1)-n}}\prod_{k=1}^{n}\frac{\Gamma(1+k\gamma)\,\Gamma(\alpha+\beta-1(n+k-2)\gamma)}{\Gamma(1+\gamma)\,\Gamma(\alpha-(k-1)\gamma)\,\Gamma(\beta-(k-1)\gamma)}$$

with $a,b,\alpha,\beta,\gamma\in\mathbb{C}$ *that satisfy*

$$\begin{aligned}&\mathfrak{Re}(z)\geqslant 0\quad\forall\, z\in\{a,b,\alpha,\beta,\gamma\}\,,\\&\mathfrak{Re}(\alpha+\beta)>1,\\&-\frac{1}{n}<\mathfrak{Re}(\gamma)<\min\left\{\frac{\mathfrak{Re}(\alpha)}{n-1},\frac{\mathfrak{Re}(\beta)}{n-1},\frac{\mathfrak{Re}(\alpha+\beta-1)}{n-1}\right\}\end{aligned}$$

The particular form of (2.6) as a product of terms is reminiscent of two things :

- a general splitting formula for Toeplitz determinants that is expressed as a product of L^2 norms of monic orthogonal polynomials for the underlying measure,
- a disintegration of the random variable $\log Z_X(0)$ as a sum of independent random variables, since its Mellin transform writes as a product of n Mellin transforms of random variables.

Toeplitz determinants

The linear statistics of the eigenangles are defined as

$$\operatorname{tr}\widetilde{f}(U_N):=\sum_{k=1}^{N}f(\alpha_k)$$

with $f:[0,2\pi]\to\mathbb{C}$ a function having a certain degree of smoothness, and $f(\theta)=\widetilde{f}(e^{i\theta})$. As one can see, the log-characteristic polynomial of a random unitary matrix $U\sim\mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$ as a function on the circle is a linear statistics.

These statistics can have their expectations rewritten in terms of a determinant. More precisely

Theorem 1.4.5 (Heine-Szegö's identity, [96]). *Let $f : [0,1] \longrightarrow \mathbb{R}$ be an integrable function and denote by $\widehat{f}(k) := \int_0^1 f(x)e^{-2i\pi kx}dx$ its k-th Fourier coefficient. Denote by*

$$D_n(f) := \det\left(\widehat{f}(k-\ell)\right)_{1\leqslant k,\ell\leqslant n}$$

the Toeplitz matrix of symbol f.

Then,

$$D_n(f) = \mathbb{E}_{\mathcal{U}(n)}\left(\prod_{k=1}^n f(\theta_k)\right) = \mathbb{E}_{\mathcal{U}(n)}\left(e^{\operatorname{tr}\log\widetilde{f}(X)}\right)$$

Proof. By Weyl's integration formula, we have

$$\mathbb{E}_{\mathcal{U}(n)}\left(e^{\operatorname{tr}\log\widetilde{f}(X)}\right) = \frac{1}{n!}\int_{[0,1]^n}\prod_{k=1}^n \widetilde{f}\left(e^{2i\pi\theta_k}\right)\left|\Delta\left(e^{2i\pi\theta_1},\dots,e^{2i\pi\theta_n}\right)\right|^2\prod_{k=1}^n d\theta_k$$

Taking linear combinations in the Vandermonde determinant, we have

$$\Delta(z_1,\dots,z_n) = \det\left(z_k^{\ell-1}\right)_{1\leqslant k,\ell\leqslant n} = \det\left(P_\ell(z_k)\right)_{1\leqslant k,\ell\leqslant n}$$

for all family of monic polynomials $(P_n)_{n\geqslant 1}$ such that $\deg(P_n) = n-1$.

Hence,

$$|\Delta(z_1,\dots,z_n)|^2 = \Delta(z_1,\dots,z_n)\overline{\Delta(z_1,\dots,z_n)} = \sum_{\sigma,\tau\in\mathfrak{S}_n}\varepsilon(\sigma\tau)\prod_{k=1}^n P_{\sigma(k)}(z_k)\overline{P_{\tau(k)}(z_k)}$$

and

$$\begin{aligned}
\int_{\mathbb{U}^n}|\Delta(z_1,\dots,z_n)|^2\prod_{k=1}^n g(z_k)dz_k &= \sum_{\sigma,\tau\in\mathfrak{S}_n}\varepsilon(\sigma\tau)\int_{\mathbb{U}^n}\prod_{k=1}^n P_{\sigma(k)}(z_k)\overline{P_{\tau(k)}(z_k)}g(z_k)dz_k \\
&= \sum_{\sigma,\tau\in\mathfrak{S}_n}\varepsilon(\sigma\tau)\prod_{k=1}^n\int_{\mathbb{U}} P_{\sigma(k)}(z)\overline{P_{\tau(k)}(z)}g(z)dz \\
&= \sum_{\sigma,\alpha\in\mathfrak{S}_n}\varepsilon(\alpha)\prod_{\ell=1}^n\int_{\mathbb{U}} P_{\alpha(\ell)}(z)\overline{P_\ell(z)}g(z)dz \\
&= n!\det\left(\int_{\mathbb{U}} P_k(z)\overline{P_\ell(z)}g(z)dz\right)_{1\leqslant k,\ell\leqslant n}
\end{aligned}$$

Taking $g(z) := \widetilde{f}(z)/(2i\pi z)$ and $P_k(z) = z^{k-1}$, we get the desired equality. $\square$

Since the equality

$$\int_{\mathbb{U}^n}|\Delta(z_1,\dots,z_n)|^2\prod_{k=1}^n g(z_k)dz_k = n!\det\left(\int_{\mathbb{U}} P_k(z)\overline{P_\ell(z)}g(z)dz\right)_{1\leqslant k,\ell\leqslant n} \tag{1.27}$$

holds for all $g \in L^1(\mathbb{U})$, if $g(z)dz$ defines a probability measure on $\mathbb{U}$, one can consider the monic orthogonal polynomials on the unit circle associated with this measure, say $(Q_n)_n$, to get

$$\int_{\mathbb{U}^n} |\Delta(z_1, \dots, z_n)|^2 \prod_{k=1}^n g(z_k) dz_k = n! \prod_{k=1}^n \|Q_k\|^2_{L^2(g \bullet \lambda_{\mathbb{U}})} \tag{1.28}$$

where $\lambda_{\mathbb{U}}$ is the Lebesgue measure on $\mathbb{U}$. Thus, dividing by $n!$, we get for instance for the modulus of the characteristic polynomial and for $s > 1$

$$\mathbb{E}_{\mathcal{U}(n)}(|Z_X(0)|^s) = \prod_{k=1}^n \|Q_k(e^{2i\pi \cdot})\|^2_{L^2\left(\left|1-e^{2i\pi\theta}\right|^s d\theta\right)}$$

The formula (1.28) explains why (2.6) takes the form of a product : the Γ terms are L^2 norms of monic orthogonal polynomials. Remark that the explicit computation of these norms gives an alternative proof of (2.6) without Selberg's integral.

One can massage the last formula in the following way

$$\begin{aligned} \|Q_k(e^{2i\pi \cdot})\|^2_{L^2\left(\left|1-e^{2i\pi\theta}\right|^s d\theta\right)} &= \int_0^1 \left|Q_k\left(e^{2i\pi\theta}\right)\right|^2 \left|1 - e^{2i\pi\theta}\right|^s d\theta \\ &= \int_0^1 \left|Q_k\left(e^{2i\pi\theta}\right)\right|^2 (2\sin(\pi\theta))^s d\theta \end{aligned}$$

Denoting by μ_k the push-forward of $\left|Q_k\left(e^{2i\pi\theta}\right)\right|^2 d\theta$ under $\theta \mapsto 2\sin(\pi\theta)$, we see that

$$\|Q_k(e^{2i\pi \cdot})\|^2_{L^2\left(\left|1-e^{2i\pi\theta}\right|^s d\theta\right)} = \int_0^1 x^s d\mu_k(x)$$

Since the polynomials Q_k depend on s, this last equality is not a priori the Mellin transform of a measure. Nevertheless, this is indeed the case.

The disintegration of the log-characteristic polynomial

Another explanation can be given to the formula (2.6). Recall that by theorem 1.2.2, we have the equality in law

$$M \stackrel{\mathcal{L}}{=} H_{u_1} H_{u_2} \dots H_{u_n}$$

where H_u is the modified Householder reflection defined in (1.10) and $u = v - e_1 \neq 0$. This last product satisfies

$$\det(I_n - H_{u_1} H_{u_2} \dots H_{u_n}) = (1 - e_1^* H_{u_1} e_1) \det(I_{n-1} - H_{u_2} \dots H_{u_n})$$

as one can see by expanding the determinant by multilinearity (for the details, see [19]). By an immediate induction

$$\det(I_n - H_{u_1} H_{u_2} \dots H_{u_n}) = \prod_{k=1}^n (1 - e_k^* H_{u_k} e_k)$$

As in theorem 1.2.2 the $(u_k)_k$ are independent random variables, this deterministic equality gives birth to an equality in law between $Z_n(0) = \det(I_n - U) \stackrel{\mathcal{L}}{=} \det(I_n - H_{u_1} H_{u_2} \dots H_{u_n})$ and a product of independent random variables. We can compute the distribution of these variables (see [19]) to get the following theorem that explicits the distribution of a Γ factor in (2.6) :

Theorem 1.4.6 (Bourgade-Hughes-Nikeghbali-Yor, [19]). *Let $U \sim \text{Haar}(\mathcal{U}_n(\mathbb{C}))$, $(U_k)_k$ a sequence of i.i.d. uniform random variables on $[0,1]$ independent of a sequence $(\beta_{1,k-1})_{1\leqslant k\leqslant n}$ of independent random variables, each defined according to the Beta distribution of parameter $(1, k-1)$, i.e. $\beta_{1,b} \stackrel{\mathcal{L}}{=} 1 - V^{1/b}$ if $V \sim \mathscr{U}([0,1])$ and $b \neq 0$ and $\beta_{1,0} \stackrel{\mathcal{L}}{=} 1 - e^{-V}$. Then, the following equality in law holds*

$$\det(I_n - U) \stackrel{\mathcal{L}}{=} \prod_{k=1}^{n} \left(1 - e^{2i\pi U_k} \sqrt{\beta_{1,k-1}}\right) \tag{1.29}$$

Proof. It is enougth to prove that $e_k^* H_{u_k} e_k \stackrel{\mathcal{L}}{=} e^{2i\pi U_k} \sqrt{\beta_{1,k-1}}$. But $H_{u_k} e_k$ in restriction to the last $n-k+1$ coordinates is a random uniform vector of $\mathbb{S}^{n-k+1}(\mathbb{C})$ and taking the scalar product with e_k, we get, with $(X_k)_k$ and $(Y_k)_k$ two independent sequences of i.i.d. Gaussian random variables

$$e_k^* H_{u_k} e_k \stackrel{\mathcal{L}}{=} \frac{X_1 + iY_1}{\sqrt{\sum_{\ell=1}^{n-k+1} (X_\ell^2 + Y_\ell^2)}} \stackrel{\mathcal{L}}{=} e^{2i\pi U_{n-k+1}} \sqrt{\beta_{1,n-k}}$$

□

One can easily check that if $Z_k := 1 - e^{2i\pi U} \sqrt{\beta_{1,k-1}}$ with $U \sim \mathscr{U}([0,1])$ independent of $\beta_{1,k-1}$, the Mellin transform of Z_k is given, for t and s such that $\mathfrak{Re}(t \pm s) > 1$ by

$$\mathbb{E}\left(e^{t \log |Z_k| + s \arg(Z_k)}\right) = \frac{\Gamma(k)\Gamma(k+t)}{\Gamma\left(k + \frac{t+is}{2}\right) \Gamma\left(k + \frac{t-is}{2}\right)}$$

which achieves to give an interpretation of (2.6).

Note that the equality (1.29) gives directly the Keating-Snaith CLT as a consequence of the classical central limit theorem of probability theory, in addition to a probabilistic interpretation of Selberg's integral in the case $\gamma = 1$, a rate of convergence by application of a classical theorem of Petrov (see [19]), etc. In particular, the speed of convergence is given by

$$\left| \mathbb{P}_{\mathcal{U}(n)}\left(\frac{\log |Z_X(0)|}{\sqrt{\frac{1}{2}\log n}} \leqslant x \right) - \int_{-\infty}^{x} e^{-u^2/2} \frac{du}{\sqrt{2\pi}} \right| \leqslant \frac{C}{(\log n)^{3/2}(1+|x|)^3}$$

where $C > 0$ is a universal constant. As in addition

$$\log |Z_n| = -\sum_{k\geqslant 1} \frac{\operatorname{tr} U^k}{k}$$

this result can be compared to the extra-exponential bound (1.14), which results of a martingale convergence with no extra renormalisation. The mixing of the traces up to a power bigger than n, from which the powers of eigenvalues are i.i.d. uniform and hence converge to the Gaussian at the classical rate of $1/\sqrt{n}$ results in a very low speed despite the fast convergence of the first terms of the series (an interesting question being to find the maximal power k_n until which the super-exponential rate is still valid).

1.4.4 The characteristic polynomial in several points of the unit circle

The splitting phenomena is thus a fundamental property of the characteristic polynomial in one point of the unit circle, but what about its behaviour in several points of the unit circle ? Following the Keating-Snaith philosophy, the answer to this question could give an idea of the behaviour of the ζ function in several points of the critical line, and in particular to a famous question about the behaviour of ratios of ζ taken in several points of a random piece of the critical line, i.e. to compute, for $U \sim \mathscr{U}([0,1])$ and integers $(b_k)_k$, $(d_k)_k$

$$\mathbb{E}\left(\prod_{k=1}^{m} \frac{\left|\zeta\left(\frac{1}{2}+it_kU\right)\right|^{b_k}}{\left|\zeta\left(\frac{1}{2}+it'_kU\right)\right|^{d_k}}\right)$$

which is a direct multivariate generalisation of the moments conjecture.

The translation of this problem in the Random Matrix world amounts to compute averages of quotients of different values of the characteristic polynomial on the unit circle, and in particular to compute

$$\mathbb{E}\left(\prod_{k=1}^{m} Z_U(\theta_k) \prod_{\ell=1}^{r} \overline{Z_U(\alpha_\ell)}\right)$$

For the general case of ratios, we refer to [26, 22]. This case leads to the

Theorem 1.4.7 (Bump-Gamburd [22], Conrey-Forrester-Snaith [26], Conrey-Farmer-Rubinstein-Keating-Snaith [25])**.** *Let* $U \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$*. Then*

$$\mathbb{E}\left(\prod_{k=1}^{m} Z_U\left(\theta_\ell\right) \prod_{\ell=1}^{r} \overline{Z_U\left(\alpha_\ell\right)}\right) = \prod_{\ell=1}^{r} e^{-ni\alpha_\ell}\, s_{\langle n^r\rangle}\left(e^{i\theta_1},\dots,e^{\theta_m},e^{\alpha_1},\dots,e^{\alpha_r}\right) \tag{1.30}$$

Proof. Setting $z_k := e^{i\theta_k}$ and $y_\ell := e^{i\alpha_\ell}$, with $Z_U\left(\theta_\ell\right) = \Phi_U\left(z_\ell\right)$, we get

$$\begin{aligned}
\mathbb{E}\left(\prod_{k=1}^{m} \Phi_U\left(z_k\right) \prod_{\ell=1}^{r} \overline{\Phi_U\left(y_\ell\right)}\right) &= \int_{\mathcal{U}_n(\mathbb{C})} \left(\prod_{k=1}^{m} \det\left(I+z_k g\right) \prod_{\ell=1}^{r} \det\left(I+y_\ell^{-1}g^{-1}\right)\right) dg \\
&= \int_{\mathcal{U}_n(\mathbb{C})} \left(\prod_{k=1}^{m} \det\left(I+z_k g\right) \prod_{\ell=1}^{r} \det\left(y_\ell^{-1}g^{-1}\right) \prod_{\ell=1}^{r} \det\left(I+y_\ell g\right)\right) dg \\
&= \prod_{\ell=1}^{r} y_\ell^{-n} \int_{\mathcal{U}_n(\mathbb{C})} \overline{\det(g)^r} \prod_{\ell=1}^{m+r} \det\left(I+z_\ell g\right) dg
\end{aligned}$$

where we have set $z_{k+\ell} = y_\ell$ for all $\ell \leqslant r$.

By the dual Cauchy identity, if $s_\lambda(g)$ represents $s_\lambda(t_1, \cdots, t_n)$ where the t_ℓ are the eigenvalues of g and s_λ the Schur function, we get

$$\begin{aligned} \prod_{\ell=1}^{m+r} \det(I + z_\ell g) &= \prod_{\ell=1}^{m+r} \prod_{j=1}^{n} (I + z_\ell t_j) = \sum_{\lambda \vdash \infty} s_\lambda(z_1, \dots, z_{m+r}) s_{\lambda'}(g) \\ \det(g)^r &= \prod_{j=1}^{n} t_j^r = s_{\langle r^n \rangle}(g) \end{aligned}$$

As $\langle n^r \rangle' = \langle r^n \rangle$, we have

$$\begin{aligned} \mathbb{E}\left(\prod_{k=1}^{r} \Phi_U(z_k) \prod_{\ell=1}^{m} \overline{\Phi_U(y_\ell)} \right) &= \prod_{\ell=1}^{r} y_\ell^{-n} \sum_{\lambda \vdash \infty} s_\lambda(z_1, \dots, z_{m+r}) \int_{\mathcal{U}_n(\mathbb{C})} \overline{s_{\langle r^n \rangle}(g)} s_{\lambda'}(g)\, dg \\ &= \prod_{\ell=1}^{r} y_\ell^{-n}\, s_{\langle n^r \rangle}(z_1, \dots, z_{m+r}) \ \text{ by orthogonality,} \end{aligned}$$

since the scalar product on the space of symmetric functions can be represented as the integral for the normalised Haar measure of $\mathcal{U}_n(\mathbb{C})$ (see [71]). This gives the result.

□

As we see, no splitting is available. Note that the integer moments of $Z_U(0)$ can be recovered by means of such a formula, giving then a combinatorial interpretation of these values (see [22] ; see also [95] for a different interpretation by means of a generalised dual Cauchy formula). Note also that other equivalent formulas are available (see [22, 26] and references cited).

1.4.5 The characteristic polynomial as a process on the unit circle and a Gaussian Field

The direct generalisation of the last computations that amounts to compute the infinite-dimensional laws of the characteristic polynomial on the unit circle is the process exhibited by Wieand in (1.17) : the *total disorder process* (see e.g. [52] for properties of this process and an arithmetic version of it).

Another generalisation of the noise hidden in the characteristic polynomial is the following : for a polynomial $P = \sum_{k=0}^{d} a_k X^k$, one can look at

$$X_n(P) := \operatorname{tr}(P(U)) = \sum_{k=0}^{d} a_k \operatorname{tr}\left(U^k\right)$$

By the asymptotic normality, we have $\operatorname{tr}(P(U)) \xrightarrow[n \to +\infty]{\mathcal{L}} \sum_{k=0}^{d} a_k \sqrt{k} Z_k$ with $(Z_k)_k$ i.i.d. complex Gaussians. And by polynomial approximation or using the Fejer kernel, one can get continuous function, the question of the general class of functions that are allowed being opened.

One possible space of functions is given by the Sobolev space of fractional derivative 1/2 defined by

$$H^{1/2} := \left\{ f \in L^2(\mathbb{U}) \ / \ \sum_{k\in\mathbb{Z}} \left|\hat{f}_k\right|^2 |k| < \infty \right\}$$

This is a Hilbert space once endowed with the scalar product

$$\langle f, g \rangle_{1/2} := \sum_{k\in\mathbb{Z}} k \hat{f}_k \overline{\hat{g}_k}$$

A probabilistic way of defining it by means of the sequence $(Z_k)_k$ of i.i.d. complex Gaussians is to say that this is the space of functions f such that

$$\left| \sum_{k\geqslant 0} \hat{f}_k \sqrt{k} Z_k \right| < \infty \ \text{ p.s.}$$

Consider the space $H_0^{1/2}$ of functions of mean 0, i.e. $\hat{f}_0 := \int_{\mathbb{U}} f = 0$. Then, we have the following convergence in law for the (centered) Gaussian field indexed by $f \in H_0^{1/2}$:

$$(X_n(f))_{f\in H_0^{1/2}} \xrightarrow[n\to+\infty]{\mathcal{L}} (X(f))_{f\in H_0^{1/2}}$$

where for all $f, g \in H_0^{1/2}$

$$\mathbb{E}\left(X(f)X(g)\right) = \langle f, g \rangle_{1/2}$$

This Gaussian field is the one naturally associated with the CUE and is obtained by a limiting procedure letting $n \to +\infty$. Nevertheless, this is only the spectrum of the matrices that is at stake, hence, one cannot speak about this construction as a $CUE(\infty)$ as in the case of virtual isometries.

Note that $z \mapsto z^k \in H_0^{1/2}$ for $k \geqslant 1$ and $|z| < 1$, so in particular, for $U \sim CUE(n)$

$$\left(\frac{\phi_U'(z)}{\phi_U(z)}\right)_{|z|<1} = \left(\sum_{k\geqslant 1} \operatorname{tr}\left(U^k\right) z^k\right)_{|z|<1} \xrightarrow[n\to+\infty]{\mathcal{L}} \left(\sum_{k\geqslant 1} \sqrt{k} Z_k z^k\right)_{|z|<1}$$

On the unit circle, the process can still be defined but with an extra renormalisation. One has, for all $\theta \notin \mathrm{Sp}(U)$

$$\log Z_n(\theta) = \log\det(I_n - e^{-i\theta}U) = -\sum_{k\geqslant 1} \operatorname{tr}\left(U^k\right) \frac{e^{-i\theta k}}{k}$$

and this process is well-defined (the principal branch of the logarithm is taken with a cut on $\mathbb{R}_-$). The Diaconis-Shashahani formula for the joint moments of the traces gives (see [33])

$$\mathbb{E}\left(\operatorname{tr} U^p \, \overline{\operatorname{tr} U^q}\right) = \mathbb{1}_{\{p=q\}} \, |p| \wedge n. \tag{1.31}$$

and the series converges in L^2 since

$$\mathbb{E}\left(|\log Z_n(\theta)|^2\right) = \sum_{k\leqslant n} \frac{1}{k} + \sum_{k\geqslant n+1} \frac{1}{k^2}$$

But as one can see, this last variance goes to $+\infty$ as $\log n$, hence the renormalisation to get the total disorder process.

1.4.6 The secular coefficients of the characteristic polynomial

There are several ways of studying the characteristic polynomial of a unitary matrix U. One of these consists in developing it in the canonical basis $\{1, X, X^2, \dots\}$, which gives

$$\Phi_U(X) := \det(I_n - XU) = (-1)^n \sum_{k=0}^{n} sc_k^{(n)} (-X)^k$$

The coefficients $(sc_k^{(n)})_k$ are the so-called *secular coefficients* of the matrix. The knowledge of their joint law is equivalent to the knowledge of the characteristic polynomial. Note that

$$\begin{aligned} sc_1^{(n)} &= \operatorname{tr}(U) \\ sc_n^{(n)} &= \det(U) \end{aligned}$$

In terms of the eigenvalues, one has moreover (e_k being the k-th elementary polynomial)

$$sc_k^{(n)} = e_k\left(e^{i\theta_{1,n}}, \dots, e^{i\theta_{n,n}}\right) := \sum_{1 \leqslant j_1 < j_2 < \cdots < j_k \leqslant n} \exp\left(i\left(\theta_{j_1,n} + \theta_{j_2,n} + \cdots + \theta_{j_k,n}\right)\right)$$

In [31], P. Diaconis and A. Gamburd make the following remark concerning the importance of these coefficients :

> A unitary matrix M is conjugate on the one hand to the diagonal matrix with eigenvalues on the diagonal, and on the other hand to the Frobenius, or companion matrix, with first row consisting of the secular coefficients, ones below the main diagonal, and remaining entries zero. This strongly suggests that secular coefficients are (as Gian-Carlo Rota might have put it) "nearly equiprimordial" with the eigenvalues.

The importance of these quantities justifies a study of their asymptotic properties, and we list here some results.

Theorem 1.4.8 (Diaconis-Gamburd, [31]). *Let $(a_1, \dots, a_\ell)$ and $(b_1, \dots, b_\ell)$ be two sequences of positive integers and let $\mu := \langle 1^{a_1}, \dots, \ell^{a_\ell}\rangle$ and $\nu := \langle 1^{b_1}, \dots, \ell^{b_\ell}\rangle$ be the associated partitions. Then,*

1. *for $n \geqslant \max\left\{\sum_j j a_j, \sum_k k b_k\right\}$, we have*

$$\mathbb{E}_{\mathcal{U}(n)}\left(\prod_{k=1}^{\ell} \left(sc_k^{(n)}\right)^{a_k} \left(\overline{sc_k^{(n)}}\right)^{b_k}\right) = N_{\mu,\nu} \tag{1.32}$$

where $N_{\mu,\nu}$ is the number of non negative integer-valued matrices A with $\operatorname{rowsum}(A) = \mu$ *and* $\operatorname{colsum}(A) = \nu$, *with, for $A = (a_{i,j})_{i,j \leqslant n}$*

$$\begin{aligned} \operatorname{rowsum}(A) &= \left(\sum_{j=1}^{n} a_{1,j}, \dots, \sum_{j=1}^{n} a_{n,j}\right) \\ \operatorname{colsum}(A) &= \left(\sum_{i=1}^{n} a_{i,1}, \dots, \sum_{i=1}^{n} a_{i,1}\right), \end{aligned}$$

2. and in particular, for $n \geqslant jk$

$$\mathbb{E}_{\mathcal{U}(n)}\left(\left|sc_j^{(n)}\right|^{2k}\right) = H_k(j) \tag{1.33}$$

where $H_k(j)$ is the number of (j,k)-magic squares, i.e. the number of $k \times k$ nonnegative integer matrices whose each row and column sum up to j.

The proof of this theorem is of the same type as the Diaconis-Shashahani theorem (1.15) and uses representation theory of $\mathcal{U}_n(\mathbb{C})$. Using (1.15), one can deduce the following limiting distribution for a fixed j

$$sc_j^{(n)} \xrightarrow[n\to+\infty]{\mathcal{L}} W_j$$

where W_j is a polynomial in the i.i.d. complex Gaussian random variables $(Z_k)_{1\leqslant k\leqslant j}$ given by

$$W_j = \frac{1}{j!}\begin{vmatrix} Z_1 & 1 & 0 & \cdots & 0 \\ \sqrt{2}\,Z_2 & Z_1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{j-1}\,Z_{j-1} & \sqrt{j-2}\,Z_{j-2} & \sqrt{j-3}\,Z_{j-3} & \cdots & j-1 \\ \sqrt{j}\,Z_j & \sqrt{j-1}\,Z_{j-1} & \sqrt{j-2}\,Z_{j-2} & \cdots & Z_1 \end{vmatrix}$$

For example, $W_3 = \frac{1}{6}Z_1^3 - \frac{1}{\sqrt{2}}Z_1Z_2 + \frac{1}{\sqrt{3}}Z_3$.

The question of a coefficient depending on n is nevertheless opened. One can remark that for $k = n$, $sc_n^{(n)} = \det(U)$ which is a random variable with values in the circle. Hence, no asymptotic normality is possible.

Another equivalent form of (1.32) and (1.33) can be given if one remarks that

$$sc_k^{(n)} = \int_0^1 \Phi_U\left(e^{2i\pi\alpha}\right) e^{-2i\pi k\alpha} d\alpha \tag{1.34}$$

This gives the

Theorem 1.4.9. *The following combinatorial description of the moments of the secular coefficients holds for all $k, j \in \mathbb{N}$*

$$\mathbb{E}_{\mathcal{U}(n)}\left(\left|sc_j^{(n)}\right|^{2k}\right) = \mathrm{card}\left\{T \in SST(\langle n^k\rangle)\ /\ \forall\, \ell \leqslant k,\ c_\ell(T) = j,\ c_{k+\ell}(T) = n-j\right\} \tag{1.35}$$

where $SST(\lambda)$ denotes the set of semi-standard Young tableaux (see e.g. [71]) for $\lambda \vdash nk$ (here $\lambda = \langle n^k\rangle$ i.e. a rectangle of height k and length n) and $c_\ell(T)$ denotes the number of times ℓ appears in T.

Proof. Using formula (1.34) and Fubini's theorem, one has

$$\begin{aligned}
\mathbb{E}_{\mathcal{U}(n)}\left(\left|sc_j^{(n)}\right|^{2k}\right) &= \mathbb{E}\left(\left|sc_j^{(n)}\right|^{2k}\right) = \mathbb{E}\left(\prod_{\ell=1}^{k} sc_j^{(n)}\overline{sc_j^{(n)}}\right) \\
&= \mathbb{E}\left(\prod_{\ell=1}^{k}\int_0^1 \Phi_U\left(e^{2i\pi\alpha_\ell}\right)e^{-2i\pi j\alpha_\ell}d\alpha_\ell \int_0^1 \overline{\Phi_U\left(e^{2i\pi\varphi_\ell}\right)e^{-2i\pi j\varphi_\ell}}d\varphi_\ell\right) \\
&= \int_{[0,1]^{2k}} \exp\left(-2i\pi j\left(\sum_{\ell=1}^{k}(\alpha_\ell-\varphi_\ell)\right)\right)\mathbb{E}\left(\prod_{\ell=1}^{k}\Phi_U\left(e^{2i\pi\alpha_\ell}\right)\overline{\Phi_U\left(e^{2i\pi\varphi_\ell}\right)}\right)d\boldsymbol{\alpha}d\boldsymbol{\varphi}
\end{aligned}$$

Using (1.30), one has

$$\mathbb{E}\left(\prod_{\ell=1}^{k}\Phi_U(z_\ell)\overline{\Phi_U(y_\ell)}\right) = \prod_{\ell=1}^{k} y_\ell^{-n}\, s_{\langle n^k\rangle}(z_1,\dots,z_k,y_1,\dots,y_k) \tag{1.36}$$

Moreover, we have the following development of Schur functions (see e.g. [71])

$$s_\lambda(x) := \sum_{T\in SST(\lambda)} x^T$$

where $x^T := \prod_{\ell\geqslant 1} x_\ell^{c_\ell(T)}$ and $c_\ell(T)$ designates the number of times ℓ appears in T.

Thus, setting $\alpha_{\ell+k}=\varphi_\ell$ in the second line,

$$\begin{aligned}
\mathbb{E}_{\mathcal{U}(n)}\left(\left|sc_j^{(n)}\right|^{2k}\right) &= \int_{[0,1]^{2k}} \exp\left(-2i\pi j\left(\sum_{\ell=1}^{k}(\alpha_\ell-\varphi_\ell)\right)\right)\mathbb{E}\left(\prod_{\ell=1}^{k}\Phi_U\left(e^{2i\pi\alpha_\ell}\right)\overline{\Phi_U\left(e^{2i\pi\varphi_\ell}\right)}\right)d\boldsymbol{\alpha}\,d\boldsymbol{\varphi} \\
&= \int_{[0,1]^{2k}} \exp\left(-2i\pi\left(\sum_{\ell=1}^{k} j(\alpha_\ell-\varphi_\ell)+n\varphi_\ell\right)\right)\sum_{T\in SST(\langle n^k\rangle)}\exp\left(2i\pi\sum_{\ell=1}^{2k}c_\ell(T)\alpha_\ell\right)d\boldsymbol{\alpha}\,d\boldsymbol{\varphi} \\
&= \sum_{T\in SST(\langle n^k\rangle)}\int_{[0,1]^{2k}}\exp\left(2i\pi\sum_{\ell=1}^{k}\alpha_\ell\left(-j+c_\ell(T)\right)+\varphi_\ell\left(j-n+c_{k+\ell}(T)\right)\right)d\boldsymbol{\alpha}\,d\boldsymbol{\varphi} \\
&= \sum_{T\in SST(\langle n^k\rangle)} \mathbb{1}_{\{\forall\,\ell\leqslant k,\, c_\ell(T)=j,\ j+c_{k+\ell}(T)=n\}} \\
&= \operatorname{card}\left\{T\in SST(\langle n^k\rangle)\ /\ \forall\,\ell\leqslant k,\, c_\ell(T)=j,\ c_{k+\ell}(T)=n-j\right\}
\end{aligned}$$

□

Remark 1.4.10. We hence have, for $n\geqslant jk$,

$$H_k(j) = \operatorname{card}\left\{T\in SST(\langle n^k\rangle)\ /\ \forall\,\ell\leqslant k,\, c_\ell(T)=j,\ c_{k+\ell}(T)=n-j\right\}$$

Since these are two cardinals of sets, an interesting combinatorial question would be to find a bijection between these sets.

Remark 1.4.11. Of course, the method also applies to compute the higher-order moments and one gets a description in terms of the SST $\langle n^{\sum_k a_k}\rangle$; nevertheless, the computations become rapidly tedious.

1.4.7 The maximum of the characteristic polynomial modulus on the unit circle

Two kinds of operations are of constant interest in probability theory : sums (linear statistics of a sequence of random variables) and maxima (extreme value statistics of a sequence of random variables). For instance, for a sequence of i.i.d. random variables, the limiting distribution (up to renormalisation) of their maxima is given by a Gumbel, a Weibull or a Fréchet distribution according to the tail of the random variables, and these limiting distributions are the only ones that can occur. Another famous extreme distribution is the Tracy-Widom one that describe the fluctuations of the largest eigenvalue of a random hermitian matrix with i.i.d. Gaussian coefficients, and that more generally appears in the context of Coulomb gases.

The extreme value statistics of the ζ function on a random uniform interval of the critical line is another famous problem in number theory (see e.g. [99]). It amounts to study the fluctuations of the random variable

$$\zeta_T^* := \max_{s\in[0,T]} \left|\zeta\left(\frac{1}{2}+isU\right)\right|$$

where $U \sim \mathscr{U}([0,1])$.

Due to the numerical scarcity of these values, no numerical simulation was able to settle the problem so far. An independent model proposed by Montgomery to predict the extreme values was based on the assumption that two local maxima of $\log|\zeta(1/2+iTU)|$ are independent (and approximately Gaussian in accordance to Selberg's CLT). This model, if true, would have implied $\log\zeta_T^*$ to be asymptotically Gumbel-distributed of order $O\left(\sqrt{\log T \log\log T}\right)$. But a straighforward computation (see [41]) shows that

$$\mathbb{E}\left(\log\left|\zeta\left(\frac{1}{2}+i(T+x_1)U\right)\right|^2 \log\left|\zeta\left(\frac{1}{2}+i(T+x_2)U\right)\right|^2\right)$$
$$\underset{T\to\infty}{\sim} \begin{cases} \log|x_1-x_2|^{-2} & \text{if} \quad \frac{1}{\log T} \ll |x_2-x_1| \ll 1 \\ 2\log\log T & \text{if} \quad |x_2-x_1| \ll \frac{1}{\log T} \end{cases}$$

This logarithmic correlation in the mid-range regime $1/\log T \ll |x_2-x_1| \ll 1$ is characteristic of a well-known Gaussian process, the $1/f$*-noise* which is the limit of the process $(\log|Z_{U_n}(\theta)|^2)_{\theta\in[0,2\pi]}$ when $n\to\infty$, for $U_n \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$. Indeed, by the Diaconis-Shashahani theorem on traces (1.15), we have for all $\ell\in\mathbb{N}$

$$\left(\frac{\mathrm{tr}(U_n^k)}{\sqrt{k}}\right)_{k\leqslant\ell} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}_{\mathbb{C}}(0, I_\ell) \tag{1.37}$$

This theorem justifies the following formal computation that was made precise in the last paragraph, with $(G_k)_k$ a sequence of i.i.d. complex Gaussian random variables :

$$\log Z_{U_n}(\theta) = \log\det\left(I_n - e^{i\theta}U_n\right) = \mathrm{tr}\log\left(I_n - e^{i\theta}U_n\right) = -\sum_{k\geqslant 1}\frac{\mathrm{tr}(U_n^k)}{k}e^{ik\theta} \underset{n\to\infty}{\approx} \sum_{k\geqslant 1}\frac{G_k}{\sqrt{k}}e^{ik\theta}$$

Taking two times the real part of this last expression gives the desired $1/f$-noise : $\log|Z_{U_n}(\theta)|$ is said to be a *regularisation* of the $1/f$-noise (see [41, 42]).

The application of the Keating-Snaith paradigm to the problem of the extreme value statistics of the Riemann Zeta function led Fyodorov and Keating to study the behaviour of the log-characteristic polynomial on the unit circle that finally led to the following conjecture :

Conjecture 1.4.12 (Fyodorov and Keating, [41])**.** The following convergence in distribution

$$\max_{\theta\in[0,2\pi]} \log|Z_{U_n}(\theta)|^2 - 2\log n + c\log\log n \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{FK}$$

holds with $c = 3/2$ and

$$\mathbb{P}(\mathcal{FK} \leqslant x) = \int_{-\infty}^{x} 2e^v K_0\left(2e^{v/2}\right) dv$$

where K_0 denotes the modified Bessel function of second kind defined for all $x > 0$ by

$$K_0(x) := \int_0^{+\infty} \frac{\cos(xt)}{\sqrt{t^2+1}} dt$$

The computations that led to conjecture 1.4.12 are only formal since they make use of a functional calculus on the $1/f$-noise, which is still unavailable, the Feynman path integral for such a noise being undefined contrary to the case of the Brownian motion where the Feynman-Kac formula would allow to do the computation. Quoting Y. V. Fyodorov (personal communication), "those predictions are well-educated conjectures based on informal but elaborate insights from theoretical physics". A first step towards their proof concerns the conjectural universal constant $c = 3/2$: it recently appeared in another instance of regularised $1/f$-noise, the $2D$ Gaussian Free Field (see [21]), giving a first credit to the conjecture since it is predicted to have a wide class of universality described by the logarithmically correlated random variables that includes in particular the 2D Gaussian free field, branching random walks, polymers on disordered trees and certain models of turbulence (see [41, 42] and references cited).

Note that conjecture 1.4.12 is equivalent to the following convergence in law

$$\frac{\max_{\theta\in[0,2\pi]} |Z_{U_n}(\theta)|}{\frac{n}{(\log n)^{3/4}}} \xrightarrow[n\to+\infty]{\mathcal{L}} Z^*$$

with

$$\mathbb{P}(Z^* \leqslant x) = \int_0^x 4K_0(2y)\, y\, dy$$

Using the Fubini theorem, we can write for all $x \geqslant 0$

$$\mathbb{P}(Z^* \leqslant x) = \int_0^{+\infty} (2tx\sin(2tx) + \cos(2tx) - 1)\frac{dt}{t^2\sqrt{1+t^2}} = 2x\int_0^{+\infty} \frac{u\sin(u)+\cos(u)-1}{u^2\sqrt{1+\left(\frac{u}{2x}\right)^2}} du$$

The Fyodorov-Keating method exposed in [41] uses a general idea of extreme value statistics : the *detropicalisation* of the functional random variable. Typically, to study

$$\|Z_U\|_\infty := \max_{\theta\in[0,2\pi]} |Z_U(\theta)|$$

one can better replace it by

$$\|Z_U\|_{2p} := \left(\int_0^1 |Z_U(2\pi\theta)|^{2p} \, d\theta \right)^{\frac{1}{2p}}$$

for $p \in \mathbb{N}^*$, since for a continuous function on $[0, 2\pi]$ as it is the case, one has the classical result

$$\|Z_U\|_{2p} \xrightarrow[p \to +\infty]{} \|Z_U\|_\infty$$

As $\|Z_U\|_{2p} \leqslant \|Z_U\|_\infty \leqslant 2^n$, one is thus led to compute

$$\phi_p(\lambda) := \mathbb{E}\left(e^{-\lambda \|Z_U\|_{2p}} \right) = \mathbb{E}\left(e^{i\lambda^{2p} \|Z_U\|_{2p}^{2p} S_{1/2p}} \right)$$

where $S_{1/2p} \sim \mathrm{SSt}\left(\frac{1}{2p}\right)$ is a random variable independent of Z_U distributed according to a symmetric stable law of parameter $\alpha = 1/2p$, i.e. for $\alpha \in]0, 2[$

$$\mathbb{E}\left(e^{itS_\alpha} \right) = e^{-|t|^\alpha}$$

We thus see that

$$\phi_p(\lambda) = \int_{\mathbb{R}_+} \mathbb{E}\left(e^{i\lambda^{2p} \|Z_U\|_{2p}^{2p} x} \right) \mathbb{P}\left(S_{1/2p} \in dx \right)$$

and the problem reduces to compute

$$\psi_p(y) := \mathbb{E}\left(e^{iy \|Z_U\|_{2p}^{2p}} \right) = \mathbb{E}\left(\exp\left(iy \int_0^1 |Z_U(2\pi\theta)|^{2p} \, d\theta \right) \right)$$

As the function $\theta \mapsto |Z_U(\theta)|$ is bounded on $[0, 2\pi]$ by 2^n, the moments are at most in geometric growth and the development in terms of moments is possible. An application of the Fubini theorem gives thus

$$\psi_p(y) = \mathbb{E}\left(\sum_{k \geqslant 0} \frac{(iy)^k}{k!} \|Z_U\|_{2p}^{2pk} \right) = \sum_{k \geqslant 0} \frac{(iy)^k}{k!} \mathbb{E}\left(\|Z_U\|_{2p}^{2pk} \right)$$

One is thus reduced to compute the following moments, for $U \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$

$$M_n^{(k)}(p) := \mathbb{E}\left(\|Z_U\|_{2p}^{2pk} \right) = \mathbb{E}\left(\left(\int_0^1 |Z_U(2\pi\theta)|^{2p} \, d\theta \right)^k \right) \tag{1.38}$$

A combinatorial description of these moments can be achieved in the same way as the moments of the coefficients of Z_U in theorem 1.4.9. For instance, for $p \in \{1, 2\}$, one has the

Theorem 1.4.13. *The following combinatorial description of $M_n^{(k)}(p)$ holds for all $k \in \mathbb{N}$ and $p = 1, 2$*

$$\begin{aligned} M_n^{(k)}(1) &= \mathrm{card}\left\{ T \in SST(\langle n^k \rangle) \;/\; \forall\, \ell \leqslant k,\, c_\ell(T) + c_{\ell+k}(T) = n \right\} \qquad (1.39) \\ M_n^{(k)}(2) &= \mathrm{card}\left\{ T \in SST(\langle n^{2k} \rangle) \;/\; \forall\, \ell \leqslant k,\, c_\ell(T) + c_{\ell+k}(T) + c_{\ell+2k}(T) + c_{\ell+3k}(T) = 2n \right\} \end{aligned}$$

where $SST(\lambda)$ and $c_\ell(T)$ are defined in theorem 1.4.9.

Proof. For $(V_\ell)_\ell$ a sequence of i.i.d. $\mathscr{U}([0,1])$-distributed random variables, write

$$M_n^{(k)}(p) := \mathbb{E}\left(\prod_{\ell=1}^k \left|\Phi_U\left(e^{2i\pi V_\ell}\right)\right|^{2p}\right) = \int_{[0,1]^k} \mathbb{E}\left(\prod_{\ell=1}^k \left|\Phi_U\left(e^{2i\pi \alpha_\ell}\right)\right|^{2p}\right) d\boldsymbol{\alpha}$$

In particular, for $p=1$, one has

$$M_n^{(k)}(1) = \int_{[0,1]^k} \mathbb{E}\left(\prod_{\ell=1}^k \left|\Phi_U\left(e^{2i\pi \alpha_\ell}\right)\right|^{2}\right) d\boldsymbol{\alpha}$$

An application of formula (1.30) gives

$$\mathbb{E}\left(\prod_{\ell=1}^k \left|\Phi_U\left(e^{2i\pi \alpha_\ell}\right)\right|^{2}\right) = \prod_{\ell=1}^k e^{-2i\pi n\alpha_\ell}\, s_{\langle n^k\rangle}\left(e^{2i\pi\alpha_1}, \dots, e^{2i\pi\alpha_k}, e^{2i\pi\alpha_1}, \dots, e^{2i\pi\alpha_k}\right)$$

which implies

$$M_n^{(k)}(1) = \int_{[0,1]^k} \exp\left(-2i\pi n \sum_{\ell=1}^k \alpha_\ell\right) s_{\langle n^k\rangle}\left(e^{2i\pi\alpha_1}, \dots, e^{2i\pi\alpha_k}, e^{2i\pi\alpha_1}, \dots, e^{2i\pi\alpha_k}\right) d\boldsymbol{\alpha}$$

Remind that Schur functions expand as

$$s_\lambda(x) = \sum_{T\in SST(\lambda)} x^T$$

with $x^T := \prod_{k\geqslant 1} x_k^{c_k(T)}$. Using this formula, we get

$$\begin{aligned} M_n^{(k)}(1) &= \int_{[0,1]^k} \exp\left(-2i\pi n \sum_{\ell=1}^k \alpha_\ell\right) \sum_{T\in SST(\langle n^k\rangle)} \exp\left(2i\pi \sum_{\ell=1}^{2k} c_\ell(T)\alpha_{\ell \bmod k}\right) d\boldsymbol{\alpha} \\ &= \sum_{T\in SST(\langle n^k\rangle)} \int_{[0,1]^k} \exp\left(2i\pi \sum_{\ell=1}^{k} \alpha_\ell\left(-n + c_\ell(T) + c_{k+\ell}(T)\right)\right) d\boldsymbol{\alpha} \\ &= \sum_{T\in SST(\langle n^k\rangle)} \mathbb{1}_{\{\forall\, \ell\leqslant k,\, c_\ell(T)+c_{k+\ell}(T)=n\}} \\ &= \operatorname{card}\left\{T \in SST(\langle n^k\rangle) \;/\; \forall\, \ell \leqslant k,\, c_\ell(T) + c_{k+\ell}(T) = n\right\} \end{aligned}$$

A straightforward application of the precedent technique gives the result for $M_n^{(k)}(2)$. □

Remark 1.4.14. The general cardinal is given by

$$M_n^{(k)}(p) = \operatorname{card}\left\{T \in SST(\langle n^{pk}\rangle) \Big/ \forall\, \ell \leqslant k,\, \sum_{r=0}^{2p-1} c_{\ell+rk}(T) = pn\right\}$$

Remark 1.4.15. The same technique of detropicalisation with stable distribution can apply to the ζ random variable, but the computation of joint moments is still conjectural and expressed by the moments conjecture.

1.4.8 The characteristic polynomial, a conclusion

The study of the Riemann ζ function is primordial for the study of arithmetic, since the Euler product development makes it a generating product of the sequence of prime numbers. As the characteristic polynomial of a random unitary matrix is a toy model for the ζ function in accordance with the Keating-Snaith philosophy, it is hence primordial to study it.

Note that the Keating-Snaith philosophy is twofold : one can use the characteristic polynomial on the circle to produce conjectures in number theory since the computations are notoriously harder to achieve in number fields, or take number-theoretic results on L-functions to check that the same results hold for characteristic polynomials on the circle. This is this last aspect that will be fully developed in the next chapter.

Chapter 2

Random matrices

If the Keating-Snaith philosophy is mainly known as a tool to produce arithmetic conjectures by means of calculations in the random matrix world (whose analogue in the number theory world seems currently out of reach), there are nonetheless certain results that can be proven on both sides, such as Selberg's central limit theorem for the Riemann zeta function and the Keating-Snaith central limit theorem for the characteristic polynomial of random unitary matrices (see [62]).

In fact Selberg's central limit theorem can be proven more generally for a wide class of L-functions (see [89] and [11]). Roughly speaking, an L-function must be defined by a Dirichlet series for $\mathfrak{Re}(s) > 1$, have an Euler product (with some growth condition on the coefficients of this product), an analytic continuation (except for finitely many poles all located on the line $\mathfrak{Re}(s) = 1$), and must satisfy a functional equation. Such L-functions are expected to satisfy the general Riemann hypothesis (GRH), which says that all the non-trivial zeros are located on the critical line, the line $\mathfrak{Re}(s) = 1/2$.

Now if one considers a finite number of such L-functions, satisfying the same functional equation, then one can wonder if the zeros of a linear combination of these L-functions are still on the critical line. The answer is in general that GRH does not hold anymore for such a linear combination even though it still has a functional equation (this can be thought of coming from the fact that such a linear combination does not have an Euler product anymore). But Bombieri and Hejhal proved in [11] that nonetheless 100% of the zeros of such linear combinations are still on the critical line (under an extra assumption of "near orthogonality" which ensures that the log of the L-functions are statistically asymptotically independent). In this chapter we will show that a similar result holds for linear combinations of independent characteristic polynomials of random unitary matrices. The result on the random matrix side is technical and difficult and besides being an extra piece of evidence that the characteristic polynomial is a good model for the value distribution of L-functions, the result is also remarkable when viewed in the general setting of random polynomials as we shall explain it. The main goal of this article is to show that on average, any linear combination of characteristic polynomials of independent random unitary matrices has a proportion of zeros on the unit circle which tends to 1 when the dimension goes to infinity.

More precisely, if U is a unitary matrix of order $N \geqslant 1$, let Φ_U be the characteristic polynomial of U, in the following sense: for $z \in \mathbb{C}$,

$$\Phi_U(z) = \det(I_N - zU).$$

From the fact that U is unitary, we get the functional equation:

$$\Phi_U(z) = (-z)^N \det(U)\overline{\Phi_U}(1/z).$$

For z on the unit circle, this equation implies that

$$\Phi_U(z) = R(z)\sqrt{(-z)^N \det(U)},$$

where $R(z)$ is real-valued (with any convention taken for the square root). The fact that Φ_U has many zeros (in fact, all of them) on the unit circle can be related to the fact that the condition needed for Φ_U to vanish is in only unidimensional (i.e. $R(z) = 0$ for a real-valued function R). Now, let $(U_j)_{1 \leqslant j \leqslant n}$ be unitary matrices of order N, and let $(b_j)_{1 \leqslant j \leqslant n}$ be real numbers: we wish to study the number of zeros on the unit circle of the linear combination

$$F_N = \sum_{j=1}^{n} b_j \Phi_{U_j}.$$

If we want that F has most of its zeros on the unit circle, it is reasonable to expect that we need a "unidimensional condition" for the equation $F(z) = 0$ if $|z| = 1$, i.e. a functional equation similar to the equation satisfied by U. This equation obviously exists if all the characteristic polynomials Φ_{U_j} satisfy the *same* functional equation, i.e. the matrices U_j have the same determinant. By symmetry of the unitary group, it is natural to assume that the unitary matrices have determinant 1. More precisely, the main result of the article is the following:

Theorem 2.0.16. *Let $(b_j)_{1 \leqslant j \leqslant n}$ be a family of (deterministic) real numbers, different from zero. For $N \geqslant 1$, let*

$$F_N := \sum_{j=1}^{n} b_j \Phi_{U_{N,j}},$$

where $(U_{N,j})_{1 \leqslant j \leqslant n}$ is a family of independent matrices following the Haar measure on the special unitary group $SU(N)$. Then, the expected proportion of zeros of F_N on the unit circle tends to 1 when N goes to infinity, i.e.

$$\mathbb{E}\left(|\{z \in \mathbb{U}, F_N(z) = 0\}|\right) = N - o(N),$$

where $|\{z \in \mathbb{U}, F_N(z) = 0\}|$ is the number of z on the unit circle which satisfy $F_N(z) = 0$.

The whole chapter is devoted to the proof of this result. Before explaining the strategy of the proof, we make a few remarks.

Remark 2.0.17. Theorem 2.0.16 can be stated as

$$\lim_{n \to \infty} \mathbb{E}\left(\frac{1}{N}|\{z \in \mathbb{U}, F_N(z) = 0\}|\right) = 1.$$

Since the random variable $\frac{1}{N}|\{z \in \mathbb{U}, F_N(z) = 0\}|$ is bounded by 1, in fact the convergence holds in all L^p spaces for $p \geqslant 1$. It also holds in probability since convergence in L^1 implies convergence in probability.

Remark 2.0.18. The fact that we impose our matrices to have the same determinant is similar to the condition in [11] of the L-functions to have the same functional equation. Moreover, in our framework, the analogue of the Riemann hypothesis is automatically satisfied since all the zeros of each characteristic polynomial are on the unit circle.

Remark 2.0.19. The fact that the proportion of zeros on the unit circle tends to 1 is a remarkable fact as a result about random polynomials. Indeed it is well known that the characteristic polynomial of a unitary matrix is self-inversive (that is $a_{N-k} = \exp(i\theta)\bar{a}_k$ for some $\theta \in \mathbb{R}$, if $(a_k)_{0 \leqslant k \leqslant N}$ are the coefficients of the polynomial). As explained in [10], self-inversive random polynomials are of interest in the context of semiclassical approximations in quantum mechanics and determining the proportion of zeros on the unit circle is there an important problem. Bogomolny, Bohigas and Leboeuf showed that if the first half of the coefficients (the second half being then fixed by the self-inverse symmetry) are chosen as independent complex Gaussian random variables, then asymptotically a fraction of $\frac{1}{\sqrt{3}}$ of the zeros are exactly on the unit circle. Hence we can say that our result is not typical of what is expected for classical random polynomials built from independent Gaussian random variables. In our framework, we do not even know the distribution of the coefficients and we also know that they are in fact not independent. Consequently the classical methods which use the independence of the coefficients (or the fact that they are Gaussian if one wants to add some dependence) would not work here. Using general results on random polynomials whose coefficients are not independent and which do not have the same distribution as stated in [51], one can deduce that the zeros cluster uniformly around the unit circle. But showing that they are almost all precisely on the unit circle is a much more refined statement.

We now say a few words about our strategy of proof of Theorem 2.0.16. In fact we use the same general method as in [11], called the "carrier waves" method, but the ingredients of our proof are different, in the sense that they are probabilistic: for instance we use the coupling method, concentration inequalities and the recent probabilistic representations of the characteristic polynomial obtained in [19]. More precisely, for $U \in U(N)$ and $t \in \mathbb{R}$, we denote by $Z_U(t)$ the characteristic polynomial of U taken at e^{-it}, i.e. $Z_U(t) = \Phi_U(e^{-it})$. Then we make a simple transformation of the linear combination F_N in order that it is real valued when restricted as a function on the unit circle:

$$i^N e^{iN\theta/2} F_N(e^{-i\theta}) = i^N e^{iN\theta/2} \sum_{j=1}^{n} b_j \Phi_{U_j}(e^{-i\theta}) = \sum_{j=1}^{n} b_j i^N e^{iN\theta/2} Z_{U_j}(\theta). \tag{2.1}$$

Using the fact that $U_j \in SU(N)$, one checks that $i^N e^{iN\theta/2} Z_{U_j}(\theta)$ is real, and that then the number of zeros of F_N on the unit circle is bounded from below by the number of sign changes, when θ increases from θ_0 to $\theta_0 + 2\pi$ (with θ_0 to be chosen carefully), of the real quantity given by the right-hand side of the equation above. The notion of carrier waves is explained in detail in [11], p. 824–827 and we do not explain it again but we would rather give a general outline. The main idea is that informally, with "high" probability and for "most" of the values of θ, one of the characteristic polynomials Z_{U_j} dominates all the others (it is the "carrier wave"). More precisely, Lemma 2.2.8 implies the following: if δ depends only on

N and tends to zero when N goes to infinity, then there exists, with probability $1-o(1)$, a subset of $[\theta_0, \theta_0+2\pi)$ with Lebesgue measure $o(1)$ such that for any θ outside this set, one can find j_0 between 1 and N such that $\log|Z_{U_{j_0}}(\theta)| - \log|Z_{U_j}(\theta)| > \delta\sqrt{\log N}$ for all $j \neq j_0$. In other words, one of the terms in the sum of the right-hand side of (2.1) should dominate all the others. Moreover, Lemma 2.2.13 informally gives the following: with high probability, the order of magnitude of each of the characteristic polynomials does not change too quickly, and then, if the interval $[\theta_0, \theta_0+2\pi)$ is divided into sufficiently many equal subintervals, the index of the carrier wave remains the same in a "large" part of each subinterval. Now, in an interval for which the carrier wave index j_0 remains the same, the zeros of $Z_{U_{j_0}}$ correspond to sign changes of $i^N e^{iN\theta/2} Z_{U_{j_0}}(\theta)$, i.e. the dominant term of (2.1). Then, one gets sign changes of $i^N e^{iN\theta/2} F_N(e^{-i\theta})$, and by counting all these sign changes, one deduces a lower bound for the number of zeros of F_N on the unit circle. The main issue of the present paper is to make rigorous this informal construction, in such a way that one gets a lower bound $N - o(N)$. One of the reasons why the proof becomes technical and involved is that we have to take into account two different kinds of sets, and show that they have almost "full measure": subsets of the interval $[\theta_0, \theta_0+2\pi)$ and subsets of $SU(N)$.

More precisely, our proof is structured as follows. We first give two standard results (Propositions 2.1.1) and 2.1.3), one on the disintegration of the Haar measure on $U(N)$ (indeed, most results on random matrices are established for $U(N)$ and we must find a way to go from the results for $U(N)$ to those for $SU(N)$) and the other one which establishes a relationship between the number of eigenvalues in a given fixed arc to the variation of the imaginary part of the log of the characteristic polynomial. Then we provide some estimates on the real and imaginary parts of the log of the characteristic polynomial (Lemmas 2.2.1 and 2.2.2) as well as a bound on the concentration of the law of the log-characteristic polynomial (Lemma 2.2.3). These estimates and some more intermediary one we establish are also useful on their own and complete the existing results in the literature on the characteristic polynomial. Then we provide bounds on the oscillations of the real and imaginary parts of the log of the characteristic polynomial (Lemma 2.2.7). We then introduce our subdivisions of the interval $[\theta_0, \theta_0+2\pi)$ and the corresponding relevant random sets to implement the carrier waves technique. Finally we combine all these estimates together to show that the average number of sign changes of (2.1) is at least $N\left(1 - O\left((\log N)^{-1/22}\right)\right)$ (the exponent $-1/22$ not being playing any major role in our analysis).

Notation

We gather here some specific notations used in this chapter.

$U(N)$ stands for the unitary group of order N, while $SU(N)$ stands for the subgroup of elements $U(N)$ whose determinant is equal to 1. $\mathbb{P}_{U(N)}$ and $\mathbb{P}_{SU(N)}$ will denote the probability Haar measure on $U(N)$ and $SU(N)$ respectively. Similarly we denote by $\mathbb{E}_{U(N)}$ and $\mathbb{E}_{SU(N)}$ the corresponding expectations.

We shall denote the Lebesgue measure on $\mathbb{R}$ by λ. If $\alpha > 0$ is a constant and if I is an interval of length α, then λ_α will denote the normalized measure $\frac{1}{\alpha}\lambda$ on the interval I.

If $\mathcal{E}$ is a finite set, we note $|\mathcal{E}|$ the number of its elements.

For n a positive integer, we note $\mathbb{P}^{(n)}_{SU(N)}$ be the n-fold product of the Haar measure on $SU(N)$, and $\mathbb{E}^{(n)}_{SU(N)}$ the corresponding expectation.

If U is a unitary matrix, remind that we denote its characteristic polynomial by $\Phi_U(z) = \det(I_N - zU)$ for $z \in \mathbb{C}$ and that for $t \in \mathbb{R}$, we denote by $Z_U(t)$ the characteristic polynomial of U taken at e^{-it}, i.e. $Z_U(t) = \Phi_U(e^{-it})$.

We shall introduce several positive quantities in the sequel: $K > 0$, $M > 0$ and $\delta > 0$. The reader should have in mind that these quantities will eventually depend on N. Unless stated otherwise, $N \geqslant 4$ and K is an integer such that $2 \leqslant K \leqslant N/2$, and $M = N/K$. In the end we will use $K \sim N/(\log N)^{3/64}$ and $\delta \sim (\log N)^{-3/32}$.

2.1 Some general facts

In this section, we state some general facts in random matrix theory, which will be used in the sequel.

2.1.1 Disintegration of the Haar measure on unitary matrices

Proposition 2.1.1. *Let $\mathbb{P}_{U(N)}$ be the Haar measure on $U(N)$, $\mathbb{P}_{SU(N)}$ the Haar measure on $SU(N)$, and for $\theta \in \mathbb{R}$, let $\mathbb{P}_{SU(N),\theta}$ be the image of $\mathbb{P}_{SU(N)}$ by the application $U \mapsto e^{i\theta}U$ from $U(N)$ to $U(N)$. Then, we have the following equality:*

$$\int_0^{2\pi} \mathbb{P}_{SU(N),\theta} \; \frac{d\theta}{2\pi} = \mathbb{P}_{U(N)}, \tag{2.2}$$

i.e. for any continuous function F from $U(N)$ to $\mathbb{R}_+$, the expectation $\mathbb{E}_{SU(N),\theta}(F)$ of F with respect to $\mathbb{P}_{SU(N),\theta}$ is measurable with respect to θ and

$$\int_0^{2\pi} \mathbb{E}_{SU(N),\theta}(F) \; \frac{d\theta}{2\pi} = \mathbb{E}_{U(N)}(F).$$

Proof. One has

$$\mathbb{E}_{SU(N),\theta}(F) = \int F(Xe^{i\theta}) d\mathbb{P}_{SU(N)}(X), \tag{2.3}$$

which, by dominated convergence, is continuous, and a fortiori measurable with respect to θ. By integrating (2.3) with respect to θ, one sees that the proposition is equivalent to the following: if U is a uniform matrix on $SU(N)$, and if Z is independent, uniform on the unit circle, then ZU is uniform on $U(N)$. Now, let A be a deterministic matrix in $U(N)$. For any $d \in \mathbb{C}$ such that $d^{-n} = \det(A)$, one has $Ad \in SU(N)$, and then $ZUA = (Z/d)(UAd)$, where:

1. UAd follows the Haar measure on $SU(N)$ (since this measure is invariant by multiplication by $Ad \in SU(N)$).

2. Z/d is uniform on the unit circle (since d, as $\det(A)$, has modulus 1).

3. These two variables, which depend deterministically on the independent variables A and Z, are independent.

Hence ZUA has the same law as ZU, i.e. this law is invariant by right-multiplication by any unitary matrix. Hence, ZU follows the Haar measure on $U(N)$. $\square$

Remark 2.1.2. This disintegration $\int_0^{2\pi} \mathbb{P}^{\otimes 2}_{e^{i\theta}SU(N)} \; \frac{d\theta}{2\pi} = \mathbb{P}^{\otimes 2}_{U(N)}$ does not hold.

2.1.2 Number of eigenvalues in an arc:

The result we state here relates the number of eigenvalues of a unitary matrix on a given arc to the logarithm of its characteristic polynomial. For $U \in U(N)$ and $t \in \mathbb{R}$, we denote by $Z_U(t)$ the characteristic polynomial of U taken at e^{-it}, i.e. $Z_U(t) = \Phi_U(e^{-it})$. Moreover, if e^{it} is not an eigenvalue of U (which occurs almost surely under Haar measure on $U(N)$, and also under the Haar measure on $SU(N)$, except for $e^{it} = 1$ and $N = 1$), we define the logarithm of $Z_U(t)$, as follows:

$$\log Z_U(t) := \sum_{j=1}^{N} \log(1 - e^{i(\theta_j - t)}), \tag{2.4}$$

where $\theta_1, \dots, \theta_N$ is the sequence of zeros of Z_U in $[0, 2\pi)$, taken with multiplicity (notice that the eigenvalues of U are $e^{i\theta_1}, \dots, e^{i\theta_N}$), and where the principal branch of the logarithm is taken in the right-hand side. We then have the following result, already stated, for example, in [49]:

Proposition 2.1.3. *Let $0 \leqslant s < t < 2\pi$, and let us assume that s and t are not zeros of Z_U. Then, the number of zeros of Z_U in the interval (s,t) is given as follows:*

$$\sum_{k=1}^{N} \mathbb{1}_{\{\theta_k \in (s,t)\}} = \frac{N}{2\pi}(t-s) + \frac{1}{\pi}\left(\mathfrak{Im}\log Z_U(t) - \mathfrak{Im}\log Z_U(s)\right). \tag{2.5}$$

Proof. It is sufficient to check that for all $\theta \in [0, 2\pi)\backslash\{s,t\}$,

$$\pi \mathbb{1}_{\{\theta \in (s,t)\}} = \frac{t-s}{2} + \mathfrak{Im}\log\left(1 - e^{i(\theta - t)}\right) - \mathfrak{Im}\log\left(1 - e^{i(\theta - s)}\right).$$

Now, for $v \in (0, 2\pi)$,

$$1 - e^{iv} = e^{iv/2}(e^{-iv/2} - e^{iv/2}) = -2i\sin(v/2)\, e^{iv/2} = 2\sin(v/2)\, e^{i(v-\pi)/2}.$$

Now, $\sin(v/2) > 0$ and $(v - \pi)/2 \in (-\pi/2, \pi/2)$ and hence

$$\mathfrak{Im}\log\left(1 - e^{iv}\right) = \frac{v - \pi}{2},$$

since we take the principal branch of the logarithm. Now, for $\theta \in [0, 2\pi)\backslash\{s,t\}$, $\theta - s + 2\pi\mathbb{1}_{\{\theta<s\}}$ and $\theta - t + 2\pi\mathbb{1}_{\{\theta<t\}}$ are in $(0, 2\pi)$, which implies

$$\begin{aligned} \mathfrak{Im}\log\left(1 - e^{i(\theta-t)}\right) - \mathfrak{Im}\log\left(1 - e^{i(\theta-s)}\right) &= \frac{\theta - t - \pi + 2\pi\mathbb{1}_{\{\theta<t\}}}{2} - \frac{\theta - s - \pi + 2\pi\mathbb{1}_{\{\theta<s\}}}{2} \\ &= \frac{s-t}{2} + \pi\left(\mathbb{1}_{\{\theta<t\}} - \mathbb{1}_{\{\theta<s\}}\right) \end{aligned}$$

and then Proposition 2.1.3.

□

2.2 Proof of the main Theorem

2.2.1 Conventions

All the random matrices we will consider are defined, for some $N \geqslant 1$, on the measurable space $(\mathcal{M}_N(\mathbb{C}), \mathcal{F})$, where $\mathcal{F}$ denotes the Borel σ-algebra of $\mathcal{M}_N(\mathbb{C})$. The canonical matrix, i.e the random variable from $(\mathcal{M}_N(\mathbb{C}), \mathcal{F})$ to $\mathcal{M}_N(\mathbb{C})$ defined by the identity function, is denoted X. Moreover, we denote by $\mathbb{E}_{U(N)}$ the expectation under $\mathbb{P}_{U(N)}$, the Haar measure on $U(N)$, and by $\mathbb{E}_{SU(N)}$ the expectation under $\mathbb{P}_{SU(N)}$, the Haar measure on $SU(N)$. For example, if F is a bounded, Borel function from $\mathcal{M}_N(\mathbb{C})$ to $\mathbb{R}$,

$$\mathbb{E}_{SU(N)}[F(X)] = \int_{\mathcal{M}_N(\mathbb{C})} F(M) d\mathbb{P}_{SU(N)}(M).$$

2.2.2 An estimate in average of the logarithm of the characteristic polynomial

Lemma 2.2.1. *There exists a universal constant $c_1 > 0$ such that for all $N \geqslant 2$, and $A \geqslant 0$,*

$$\int_0^{2\pi} \mathbb{P}_{SU(N)}\left(\left|\log Z_X(\theta)\right| \geqslant A\sqrt{\log N}\right) \frac{d\theta}{2\pi} \leqslant c_1 e^{-\frac{A}{2}\left(A \wedge \frac{\sqrt{\log N}}{2}\right)}$$

Proof. For all $\lambda \geqslant 0$,

$$\begin{aligned}
\int_0^{2\pi} \mathbb{P}_{SU(N)}\left(\left|\log Z_X(\theta)\right| \geqslant A\sqrt{\log N}\right) \frac{d\theta}{2\pi} &\leqslant e^{-\lambda A\sqrt{\log N}} \int_0^{2\pi} \mathbb{E}_{SU(N)}\left(e^{\lambda|\log Z_X(\theta)|}\right) \frac{d\theta}{2\pi} \\
&\leqslant e^{-\lambda A\sqrt{\log N}} \mathbb{E}_{U(N)}\left(e^{\lambda|\log Z_X(0)|}\right) \text{ (by (2.2))} \\
&\leqslant e^{-\lambda A\sqrt{\log N}} \mathbb{E}_{U(N)}\left(e^{\lambda(|\mathfrak{Re}\log Z_X(0)|+|\mathfrak{Im}\log Z_X(0)|)}\right)
\end{aligned}$$

Using the inequality $e^{|a|+|b|} \leqslant e^{a+b}+e^{a-b}+e^{-a+b}+e^{-a-b}$, valid for all $a, b \in \mathbb{R}$, and writing the right-hand side of this inequality as $4\mathbb{E}\left(e^{Ba+B'b}\right)$ for B and B' being two independent Bernoulli random variables independent of U such that $\mathbb{P}(B=1) = 1-\mathbb{P}(B=-1) = 1/2$, we have:

$$\int_0^{2\pi} \mathbb{P}_{SU(N)}\left(\left|\log Z_X(\theta)\right| \geqslant A\sqrt{\log N}\right) \frac{d\theta}{2\pi} \leqslant 4e^{-\lambda A\sqrt{\log N}} \mathbb{E}_{U(N)}\left(e^{\lambda(B\,\mathfrak{Re}\log Z_X(0)+B'\mathfrak{Im}\log Z_X(0))}\right).$$

We now use the fact ([62] and [48], p.16) that for $s, t \in \mathbb{C}$ such that $\mathfrak{Re}(s+it)$ and $\mathfrak{Re}(s-it)$ are strictly larger than -1:

$$\mathbb{E}_{U(N)}\left(e^{s\,\mathfrak{Re}\log Z_X(0)+t\,\mathfrak{Im}\log Z_X(0)}\right) = \frac{G\left(1+\frac{s+it}{2}\right)G\left(1+\frac{s-it}{2}\right)G(1+N)G(1+N+s)}{G\left(1+N+\frac{s+it}{2}\right)G\left(1+N+\frac{s-it}{2}\right)G(1+s)} \tag{2.6}$$

where G is the Barnes G-function, defined for all $z \in \mathbb{C}$, by

$$G(z+1) := (2\pi)^{z/2} e^{-[(1+\gamma)z^2+z]/2} \prod_{n=1}^{\infty} \left(1+\frac{z}{n}\right)^n e^{-z+(z^2/2n)},$$

γ being the Euler constant. The function G also satisfies the functional equation $G(z+1) = \Gamma(z)G(z)$.

In other words, one has

$$\mathbb{E}_{U(N)}\left(e^{s\,\mathfrak{Re}\log Z_X(0)+t\,\mathfrak{Im}\log Z_X(0)}\right) = \frac{G\left(1+\frac{s+it}{2}\right)G\left(1+\frac{s-it}{2}\right)}{G(1+s)} N^{\left(s^2+t^2\right)/4} G_{N,s,t},$$

where, by the classical estimates of the Barnes function,

$$G_{N,s,t} := N^{-\left(s^2+t^2\right)/4} \frac{G(1+N)\,G(1+N+s)}{G\left(1+N+\frac{s+it}{2}\right)G\left(1+N+\frac{s-it}{2}\right)}$$

tends to 1 when N goes to infinity, uniformly on s and t if these parameters are bounded.

For any sequence $(\lambda_N)_{N\geqslant 1}$ such that $\lambda_N \in [0,1/2]$, one has (taking $s = \lambda_N B$ and $t = \lambda_N B'$):

$$\mathbb{E}_{U(N)}\left(e^{\lambda_N(B\,\mathfrak{Re}\log Z_X(0)+B'\mathfrak{Im}\log Z_X(0))}\right) = M(\lambda_N) N^{\frac{\lambda_N^2}{2}},$$

where

$$M(\lambda_N) := \mathbb{E}\left(\frac{G\left(1+\lambda_N\frac{B+iB'}{2}\right)G\left(1+\lambda_N\frac{B-iB'}{2}\right)}{G(1+\lambda_N B)} G_{N,\lambda_N B,\lambda_N B'}\right).$$

Since the function G is holomorphic, with no zero on the half-plane $\{\mathfrak{Re} > 0\}$, and since $G_{N,\lambda B,\lambda B'}$ tends to 1 when N goes to infinity, uniformly on $\lambda \in [0,1/2]$, the quantity $M(\lambda)$ is uniformly bounded by some universal constant $c' > 0$, for $\lambda \in [0,1/2]$. Hence,

$$\mathbb{E}_{U(N)}\left(e^{\lambda_N(B\,\mathfrak{Re}\log Z_X(0)+B'\mathfrak{Im}\log Z_X(0))}\right) \leqslant c' N^{\frac{\lambda_N^2}{2}},$$

for N going to infinity, which implies:

$$\int_0^{2\pi} \mathbb{P}_{SU(N)}\left(\left|\log Z_X(\theta)\right| \geqslant A\sqrt{\log N}\right)\frac{d\theta}{2\pi} \leqslant 4c' e^{-\lambda_N A\sqrt{\log N}+(\lambda_N^2\log N)/2}.$$

Now, taking $\lambda_N = (1/2)\wedge(A/\sqrt{\log N})$ gives

$$\begin{aligned}
\int_0^{2\pi} \mathbb{P}_{SU(N)}\left(\left|\log Z_X(\theta)\right| \geqslant A\sqrt{\log N}\right)\frac{d\theta}{2\pi} &\leqslant 4c' e^{-\lambda_N\sqrt{\log N}[A-(\lambda_N\sqrt{\log N})/2]} \\
&\leqslant 4c' e^{-\lambda_N\sqrt{\log N}[A-(A/\sqrt{\log N})(\sqrt{\log N})/2]} \\
&\leqslant 4c' e^{-\lambda_N\sqrt{\log N}(A/2)} \\
&\leqslant 4c' e^{-[(\sqrt{\log N}/2)\wedge A](A/2)}
\end{aligned}$$

$\square$

2.2.3 An estimate on the imaginary part of the log-characteristic polynomial

From the previous result, we obtain the following estimate for the imaginary part of the log-characteristic polynomial:

Lemma 2.2.2. *There exists a universal constant $c_1' > 0$ such that for all $N \geqslant 2$, $A \geqslant 0$, and $\theta \in \mathbb{R}$,*

$$\mathbb{P}_{SU(N)}\left(\left|\mathfrak{Im}\log Z_X(\theta)\right| \geqslant A\sqrt{\log N}\right) \leqslant c_1' e^{-\frac{A}{2}\left(A\wedge\frac{\sqrt{\log N}}{2}\right)}$$

Proof. We use here the probabilistic splitting established in [19] which shows that (see also [14] for an infinite-dimensional point of view), for any $U \in U(N)$, there exists, for $1 \leqslant j \leqslant N$, x_j on the unit sphere of $\mathbb{C}^j$, uniquely determined, such that

$$U = R(x_N)\begin{pmatrix} R(x_{N-1}) & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} R(x_{N-2}) & 0 \\ 0 & I_2 \end{pmatrix}\cdots\begin{pmatrix} R(x_1) & 0 \\ 0 & I_{N-1} \end{pmatrix}, \tag{2.7}$$

where $R(x_j)$ denotes the unique unitary matrix in $U(j)$ sending the last basis vector e_j of $\mathbb{C}^j$ to x_j, and such that the image of $I_j - R(x_j)$ is the vector space generated by $e_j - x_j$.

Moreover, the characteristic polynomial of $U(N)$ is given by

$$Z_U(0) = \prod_{j=1}^{N}(1 - \langle x_j, e_j\rangle),$$

and then its logarithm is

$$\log Z_U(0) = \sum_{j=1}^{N}\log(1 - \langle x_j, e_j\rangle), \tag{2.8}$$

when 1 is not an eigenvalue of U, taking the principal branch of the logarithm on the right-hand side. Notice that the determination of the logarithm given by this formula fits with the definition involving the eigenangles (2.4). Indeed, the two formulas depend continuously on the matrix U, on the connected set $\{U \in U(N), 1 \notin \mathrm{Spec}(U)\}$, and their exponentials are equal, hence, it is sufficient to check that they coincide for one matrix U, for example $-I_N$ (in this case, $x_j = -e_j$ for all j and the two formulas give $N\log 2$).

If U follows the uniform distribution on $U(N)$, then the vectors $(x_j)_{1\leqslant j\leqslant N}$ are independent and x_j is uniform on the sphere of $\mathbb{C}^j$. The determinant of U is equal to the product of the determinants of $R(x_j)$ for $1 \leqslant j \leqslant N$, and since $R(x_1)$ is the multiplication by x_1 on $\mathbb{C}$, one has

$$\det(U) = x_1\prod_{j=2}^{N}\Gamma_j(x_j),$$

where Γ_j is a function from $\mathbb{C}^j$ to the unit circle $\mathbb{U}$. From this, let us deduce that under the measure $\mathbb{P}_{SU(N),\theta}$:

1. The vectors $(x_j)_{2\leqslant j\leqslant N}$ are independent, x_j being uniform on the unit sphere of $\mathbb{C}^j$.

2. The value of $x_1 \in \mathbb{U}$ is uniquely determined by the determinant $\det(U) = e^{iN\theta}$,

$$x_1 = e^{iN\theta}\prod_{j=2}^{N}[\Gamma_j(x_j)]^{-1}.$$

Indeed, let $\mathbb{P}'_{SU(N),\theta}$ be the probability measure on the image of $SU(N)$ by the multiplication by $e^{i\theta}$, under which the law of $(x_j)_{1\leqslant j\leqslant N}$ is given by the two items above. This probability measure can be constructed as the law of the random matrix U given by the formula (2.7), where $(x_j)_{1\leqslant j\leqslant N}$ are random vectors whose joint distribution is given by the items (1) and (2) just above. We now have to prove that $\mathbb{P}_{SU(N),\theta} = \mathbb{P}'_{SU(N),\theta}$. Let us first notice that the joint law of $(x_j)_{2\leqslant j\leqslant N}$, under the probability measure $\mathbb{P}'_{SU(N),\theta}$, does not depend on θ. Hence, under the averaged measure

$$\int_0^{2\pi} \mathbb{P}'_{SU(N),\theta} \frac{d\theta}{2\pi},$$

the vectors $(x_j)_{2\leqslant j\leqslant N}$ still have the same law, i.e. they are independent and x_j is uniform on the unit sphere of $\mathbb{C}^j$. Moreover, conditionally on $(x_j)_{2\leqslant j\leqslant N}$, $x_1 = e^{iN\theta}\prod_{j=2}^N [\Gamma_j(x_j)]^{-1}$, where θ is uniform on $[0, 2\pi)$. Hence, $(x_j)_{1\leqslant j\leqslant N}$ are independent, x_1 is uniform on $\mathbb{U}$, and then x_j in uniform on the unit sphere of $\mathbb{C}^j$ for all $j \in \{1,\dots,N\}$, which implies

$$\int_0^{2\pi} \mathbb{P}'_{SU(N),\theta} \frac{d\theta}{2\pi} = \mathbb{P}_{U(N)} = \int_0^{2\pi} \mathbb{P}_{SU(N),\theta} \frac{d\theta}{2\pi}.$$

Now, $\mathbb{P}_{SU(N),2\pi/N}$ is the image of $\mathbb{P}_{SU(N)}$ by multiplication by $e^{i2\pi/N} I_N$, which is a matrix in $SU(N)$: the invariance property defining the Haar measure $\mathbb{P}_{SU(N)}$ implies that $\mathbb{P}_{SU(N),2\pi/N} = \mathbb{P}_{SU(N)}$, and then $\theta \mapsto \mathbb{P}_{SU(N),\theta}$ is $(2\pi/N)$-periodic. It is the same for $\theta \mapsto \mathbb{P}'_{SU(N),\theta}$, since the values of $x_1, \dots x_N$ involved in the definition of $\mathbb{P}'_{SU(N),\theta}$ do not change if we add a multiple of $2\pi/N$ to θ. Hence,

$$\int_0^{2\pi/N} \mathbb{P}'_{SU(N),\theta} \frac{N d\theta}{2\pi} = \int_0^{2\pi/N} \mathbb{P}_{SU(N),\theta} \frac{N d\theta}{2\pi}.$$

Now, let F be a continuous, bounded function from $U(N)$ to $\mathbb{R}$. By applying the equality above to the function $U \mapsto F(U)\mathbb{1}_{\{\det U \in \{e^{iN\theta},\, \theta \in I\}\}}$, for an interval $I \subset [0, 2\pi/N)$, one deduces with obvious notation that:

$$\int_I \mathbb{E}'_{SU(N),\theta}(F) \frac{d\theta}{|I|} = \int_I \mathbb{E}_{SU(N),\theta}(F) \frac{d\theta}{|I|},$$

where $|I|$ is the length of I. Now, by definition of $\mathbb{P}_{SU(N),\theta}$ and $\mathbb{P}'_{SU(N),\theta}$, the first measure is the image of $\mathbb{P}_{SU(N)}$ by multiplication by $e^{i\theta}$, and the second measure is the image of $\mathbb{P}'_{SU(N),0}$ by right multiplication by the matrix $\begin{pmatrix} e^{iN\theta} & 0 \\ 0 & I_{N-1} \end{pmatrix}$. Hence, by continuity and boundedness of F, and by dominated convergence, $\mathbb{E}_{SU(N),\theta}(F)$ and $\mathbb{E}'_{SU(N),\theta}(F)$ are continuous with respect to θ. By considering a sequence $(I_r)_{r\geqslant 1}$ of intervals containing a given value of θ and whose length tends to zero, one deduces, by letting $r \to \infty$,

$$\mathbb{E}'_{SU(N),\theta}(F) = \mathbb{E}_{SU(N),\theta}(F).$$

We now get the equality $\mathbb{P}_{SU(N),\theta} = \mathbb{P}'_{SU(N),\theta}$, and then the law of $(x_j)_{1\leqslant j\leqslant N}$ under $\mathbb{P}_{SU(N),\theta}$ described above.

Hence, the sequence $(x_j)_{2\leqslant j\leqslant N}$ has the same law under $\mathbb{P}_{SU(N),\theta}$ and $\mathbb{P}_{U(N)}$. We now use this fact to construct a coupling between these two probability measures on the unitary group.

The general principle of coupling is the following: when we want to show that two probability distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ on a metric space have a similar behavior, a possible strategy is to construct a couple (U, U') of random variables defined on the same probability space endowed with a probability $\mathbb{P}$, such that the law of U under $\mathbb{P}$ is $\mathbb{P}_1$, the law of U' under $\mathbb{P}$ is $\mathbb{P}_2$, and the distance between U and U' is small with high probability. In the present situation, we take $(x'_j)_{1\leqslant j\leqslant N}$ independent, x'_j uniform on the unit sphere of $\mathbb{C}^j$ for all $j \in \{1,\dots,N\}$, and we construct, by using (2.7), a random matrix U' following $\mathbb{P}_{U(N)}$. Then, we do the coupling by taking $x_j := x'_j$ for $2 \leqslant j \leqslant N$ and

$$x_1 := e^{iN\theta}\prod_{j=2}^{N}[\Gamma_j(x_j)]^{-1},$$

which gives a random matrix U following $\mathbb{P}_{SU(N),\theta}$. From the fact that $x_j = x'_j$ for $j \geqslant 2$ and the equation (2.8), we get the following:

$$\log Z_U(0) - \log Z_{U'}(0) = \log(1-x_1) - \log(1-x'_1),$$

and in particular,

$$|\mathfrak{Im}\log Z_U(0) - \mathfrak{Im}\log Z_{U'}(0)| \leqslant \pi.$$

Now, for $B := \left(A - \frac{\pi}{\sqrt{\log N}}\right)_+$, one gets :

$$\begin{aligned}
\mathbb{P}_{SU(N)}\left(\left|\mathfrak{Im}\log Z_X(-\theta)\right| \geqslant A\sqrt{\log N}\right) &= \mathbb{P}_{SU(N),\theta}\left(\left|\mathfrak{Im}\log Z_X(0)\right| \geqslant A\sqrt{\log N}\right)\\
&= \mathbb{P}\left(\left|\mathfrak{Im}\log Z_U(0)\right| \geqslant A\sqrt{\log N}\right)\\
&\leqslant \mathbb{P}\left(\left|\mathfrak{Im}\log Z_{U'}(0)\right| \geqslant A\sqrt{\log N} - \pi\right)\\
&= \mathbb{P}_{U(N)}\left(\left|\mathfrak{Im}\log Z_X(0)\right| \geqslant B\sqrt{\log N}\right)\\
&= \int_0^{2\pi}\mathbb{P}_{SU(N)}\left(\left|\mathfrak{Im}\log Z_X(\theta)\right| \geqslant B\sqrt{\log N}\right)\frac{d\theta}{2\pi}\\
&\leqslant \int_0^{2\pi}\mathbb{P}_{SU(N)}\left(\left|\log Z_X(\theta)\right| \geqslant B\sqrt{\log N}\right)\frac{d\theta}{2\pi}\\
&\leqslant c_1 e^{-\frac{B}{2}\left(B\wedge\frac{\sqrt{\log N}}{2}\right)}
\end{aligned}$$

Now, if $B \leqslant \frac{\sqrt{\log N}}{2}$,

$$\begin{aligned}
\frac{A}{2}\left(A\wedge\frac{\sqrt{\log N}}{2}\right) \leqslant \frac{A^2}{2} \leqslant \frac{1}{2}\left(B + \frac{\pi}{\sqrt{\log N}}\right)^2 &= \frac{B^2}{2} + \frac{B\pi}{\sqrt{\log N}} + \frac{\pi^2}{2\log N}\\
&\leqslant \frac{B^2}{2} + \frac{\pi}{2} + \frac{\pi^2}{2\log 2}\\
&= \frac{B}{2}\left(B\wedge\frac{\sqrt{\log N}}{2}\right) + \frac{\pi}{2} + \frac{\pi^2}{2\log 2}
\end{aligned}$$

If $B \geqslant \frac{\sqrt{\log N}}{2}$,

$$\frac{A}{2}\left(A \wedge \frac{\sqrt{\log N}}{2}\right) \leqslant \frac{A\sqrt{\log N}}{4} \leqslant \frac{\sqrt{\log N}}{4}\left(B + \frac{\pi}{\sqrt{\log N}}\right) = \frac{B\sqrt{\log N}}{4} + \frac{\pi}{4}$$
$$= \frac{B}{2}\left(B \wedge \frac{\sqrt{\log N}}{2}\right) + \frac{\pi}{4}$$

Hence, we get Lemma 2.2.2, with

$$c_1' = c_1 e^{\frac{\pi}{2} + \frac{\pi^2}{2\log 2}}.$$

□

2.2.4 Bound on the concentration of the law of the log-characteristic polynomial

Lemma 2.2.3. *For $N \geqslant 4$, $\theta \in [0, 2\pi)$, $x_0 \in \mathbb{R}$ and $\delta \in (0, 1/2)$, one has*

$$\mathbb{P}_{SU(N)}[|\log|Z_X(\theta)| - x_0| \leqslant \delta\sqrt{\log N}] \leqslant C\delta\log(1/\delta),$$

where $C > 0$ is a universal constant.

Proof. The proof of Lemma 2.2.3 needs several steps.

Sublemma 2.2.4. *For $j \geqslant 1$ integer, $s, t \in \mathbb{R}$, let us define*

$$Q(j, s, t) := \frac{\left(j + \frac{it-s}{2}\right)\left(j + \frac{it+s}{2}\right)}{j(j+it)}.$$

Then,

1. *For $s^2 + t^2 \geqslant 8j^2$, $|Q(j, s, t)| \geqslant \max\left(1, \frac{\sqrt{s^2+t^2}}{8j}\right)$.*
2. *For $j^2 \leqslant s^2 + t^2 \leqslant 8j^2$, $|Q(j, s, t)| \leqslant 1$.*
3. *For $s^2 + t^2 \leqslant j^2$, $|Q(j, s, t)| \leqslant e^{-(s^2+t^2)/10j^2}$.*

Proof. One has:

$$Q(j, s, t) = \frac{1 - \frac{s^2+t^2}{4j^2} + it/j}{1 + it/j}. \tag{2.9}$$

If $s^2 + t^2 \leqslant 8j^2$, it is immediate that the numerator has a smaller absolute value than the denominator, i.e. $|Q(j, s, t)| \leqslant 1$. Moreover,

$$|Q(j, s, t)|^2 = \frac{1 - \frac{s^2+t^2}{2j^2} + \frac{(s^2+t^2)^2}{16j^4} + \frac{t^2}{j^2}}{1 + \frac{t^2}{j^2}} = 1 - \frac{\left(\frac{s^2+t^2}{2j^2}\right)\left(1 - \frac{s^2+t^2}{8j^2}\right)}{1 + \frac{t^2}{j^2}}$$

and in the case where $s^2 + t^2 \leqslant j^2$, one deduces

$$|Q(j, s, t)|^2 \leqslant 1 - \frac{7(s^2+t^2)}{32j^2}$$

and then

$$|Q(j,s,t)| \leqslant e^{-7(s^2+t^2)/64j^2} \leqslant e^{-(s^2+t^2)/10j^2}.$$

Now, if $s^2+t^2 \geqslant 8j^2$, the numerator in (2.9) has a larger absolute value than the denominator, and then $|Q(j,s,t)| \geqslant 1$. Moreover, since $(s^2+t^2)/8j^2 \geqslant 1$,

$$\begin{aligned}
|Q(j,s,t)|^2 &= \frac{\left(\frac{s^2+t^2}{4j^2}-1\right)^2+\frac{t^2}{j^2}}{1+\frac{t^2}{j^2}} \geqslant \frac{\left(\frac{s^2+t^2}{8j^2}\right)^2+\frac{t^2}{j^2}}{1+\frac{t^2}{j^2}} \\
&\geqslant \frac{\left(\frac{s^2+t^2}{8j^2}\right)^2+\frac{s^2+t^2}{j^2}}{1+\frac{s^2+t^2}{j^2}} \\
&\geqslant \frac{1}{64}\cdot\frac{\left(\frac{s^2+t^2}{j^2}\right)^2+\frac{s^2+t^2}{j^2}}{1+\frac{s^2+t^2}{j^2}} = \frac{s^2+t^2}{64j^2}
\end{aligned}$$

which finishes the proof of the sublemma.

□

Sublemma 2.2.5. *Let $j \geqslant 1$ be an integer, let ρ_j and σ_j be the real and imaginary parts of $\log(1-\sqrt{\beta_{1,j-1}}e^{i\theta})$, where $\beta_{1,j-1}$ is a beta random variable with $\beta(1,j-1)$ distribution and θ is independent of $\beta_{1,j-1}$, uniform on $[0,2\pi]$. Then, for $s,t\in\mathbb{R}$,*

$$|\mathbb{E}[e^{i(t\rho_j+s\sigma_j)}]| \leqslant e^{-(s^2+t^2)/30j}$$

if $s^2+t^2 \leqslant 8j^2$, and

$$|\mathbb{E}[e^{i(t\rho_j+s\sigma_j)}]| \leqslant \frac{8}{\sqrt{s^2+t^2}}$$

if $s^2+t^2 \geqslant 8j^2$ and $j \geqslant 2$.

Proof. For $t\in\mathbb{R}$ and $s\in\mathbb{C}$ with real part strictly between -1 and 1,

$$\mathbb{E}[e^{i(t\rho_j+s\sigma_j)}] = \frac{\Gamma(j)\Gamma(j+it)}{\Gamma\left(j+\frac{it-s}{2}\right)\Gamma\left(j+\frac{it+s}{2}\right)} \tag{2.10}$$

(see [19]). Now, if t is fixed, the function

$$s \mapsto \mathbb{E}[e^{i(t\rho_j+s\sigma_j)}]$$

is holomorphic, since the imaginary part is uniformly bounded (by $\pi/2$), which implies that (2.10) holds for all $t\in\mathbb{R}$, $s\in\mathbb{C}$, and in particular for all $s,t\in\mathbb{R}$. Moreover,

$$\frac{\Gamma(k)\Gamma(k+it)}{\Gamma\left(k+\frac{it-s}{2}\right)\Gamma\left(k+\frac{it+s}{2}\right)} \underset{k\to\infty}{\longrightarrow} 1,$$

since $\Gamma(k+z)/\Gamma(k)$ is equivalent to k^z for all $z\in\mathbb{C}$. Hence, by using the equation $\Gamma(z+1) = z\Gamma(z)$, one deduces:

$$\mathbb{E}[e^{i(t\rho_j+s\sigma_j)}] = \prod_{k=j}^{\infty}\frac{\left(k+\frac{it-s}{2}\right)\left(k+\frac{it+s}{2}\right)}{k(k+it)} = \prod_{k=j}^{\infty} Q(k,s,t).$$

If $s^2+t^2 \leqslant 8j^2$, then $|Q(k,s,t)| \leqslant 1$ for all $k \geqslant j$ and $|Q(k,s,t)| \leqslant e^{-(s^2+t^2)/10k^2}$ for all $k \geqslant 3j$. Hence

$$|\mathbb{E}[e^{i(t\rho_j+s\sigma_j)}]| \leqslant \prod_{k=3j}^{\infty} e^{-(s^2+t^2)/10k^2} \leqslant \prod_{k=3j}^{\infty} e^{-(s^2+t^2)/10k(k+1)} = e^{-(s^2+t^2)/30j}.$$

Now let us assume $s^2+t^2 \geqslant 8j^2$. One has:

$$\mathbb{E}[e^{i(t\rho_j+s\sigma_j)}] = \frac{\Gamma(1)\Gamma(1+it)}{\Gamma\left(1+\frac{it-s}{2}\right)\Gamma\left(1+\frac{it+s}{2}\right)} \prod_{k=1}^{j-1} \frac{1}{Q(k,s,t)}$$

where all the factors $\frac{1}{Q(k,s,t)}$ have absolute value bounded by one. By considering the case where $j=1$, one deduces

$$\left|\frac{\Gamma(1)\Gamma(1+it)}{\Gamma\left(1+\frac{it-s}{2}\right)\Gamma\left(1+\frac{it+s}{2}\right)}\right| \leqslant 1,$$

and then, for $j \geqslant 2$,

$$|\mathbb{E}[e^{i(t\rho_j+s\sigma_j)}]| \leqslant \frac{1}{|Q(1,s,t)|} \leqslant \frac{8}{\sqrt{s^2+t^2}}.$$

□

Sublemma 2.2.6. *For $N \geqslant 4$ and $\theta \in [0,2\pi)$, the distribution of $\log(Z_X(\theta))$ under Haar measure on $U(N)$ has a density with respect to Lebesgue measure on $\mathbb{C}$, which is continuous and bounded by $C_0/\log(N)$, where $C_0 > 0$ is a universal constant.*

Proof. By the results in [19] and the previous sublemma, one checks that the characteristic function Φ of $\log(Z_X(\theta)) \in \mathbb{C} \sim \mathbb{R}^2$ is given by

$$\Phi(s,t) = \prod_{j=1}^{N} \mathbb{E}[e^{i(t\rho_j+s\sigma_j)}].$$

If $s^2+t^2 \geqslant 32N$, one has $s^2+t^2 \geqslant 128 \geqslant 8j^2$ for $j \in \{2,3,4\}$. Hence,

$$|\Phi(s,t)| \leqslant |\mathbb{E}[e^{i(t\rho_2+s\sigma_2)}]||\mathbb{E}[e^{i(t\rho_3+s\sigma_3)}]||\mathbb{E}[e^{i(t\rho_4+s\sigma_4)}]| \leqslant \frac{512}{(s^2+t^2)^{3/2}}.$$

If $s^2+t^2 \leqslant 32N$, then $s^2+t^2 \leqslant 8j^2$ for all $j \geqslant 2\sqrt{N}$. Hence,

$$|\Phi(s,t)| \leqslant \prod_{2\sqrt{N}\leqslant j\leqslant N} \mathbb{E}[e^{i(t\rho_j+s\sigma_j)}] \leqslant \exp\left(-(s^2+t^2)\sum_{2\sqrt{N}\leqslant j\leqslant N} \frac{1}{30j}\right).$$

Since $e^{1/j} \geqslant \frac{j+1}{j}$, one deduces

$$\begin{aligned}|\Phi(s,t)| &\leqslant \prod_{2\sqrt{N}\leqslant j\leqslant N} \left(\frac{j}{j+1}\right)^{(s^2+t^2)/30} \leqslant \left(\frac{2\sqrt{N}+1}{N+1}\right)^{(s^2+t^2)/30} \\ &\leqslant \left(\frac{3\sqrt{N}}{N}\right)^{(s^2+t^2)/30} = e^{-\log(N/9)(s^2+t^2)/60}\end{aligned}$$

Now, for $N \geqslant 10$,

$$\begin{aligned}
\int_{\mathbb{R}^2} |\Phi(s,t)| ds dt &\leqslant \int_{\mathbb{R}^2} \frac{512}{(s^2+t^2)^{3/2}} \mathbb{1}_{\{s^2+t^2 \geqslant 32N\}} ds dt + \int_{\mathbb{R}^2} e^{-\log(N/9)(s^2+t^2)/60} \mathbb{1}_{\{s^2+t^2 \leqslant 32N\}} ds dt \\
&= \pi \left(\int_0^{32N} e^{-u \log(N/9)/60} du + \int_{32N}^{\infty} \frac{512}{u^{-3/2}} du \right) \\
&\leqslant \frac{60\pi}{\log(N/9)} + 1024\pi (32N)^{-1/2} \leqslant \frac{10000}{\log N},
\end{aligned}$$

and for $N \in \{4,5,6,7,8,9\}$,

$$\begin{aligned}
\int_{\mathbb{R}^2} |\Phi(s,t)| ds dt &\leqslant \int_{\mathbb{R}^2} \frac{512}{(s^2+t^2)^{3/2}} \mathbb{1}_{\{s^2+t^2 \geqslant 32N\}} ds dt + \int_{\mathbb{R}^2} \mathbb{1}_{\{s^2+t^2 \leqslant 32N\}} ds dt \\
&= \pi \left(\int_0^{32N} du + \int_{32N}^{\infty} \frac{512}{u^{-3/2}} du \right) \\
&\leqslant 32\pi N + 1024\pi (32N)^{-1/2} \leqslant 288\pi + 1024\pi (128)^{-1/2} \leqslant \frac{10000}{\log 9}.
\end{aligned}$$

By applying Fourier inversion, we obtain Sublemma 2.2.6.

□

Let us now go back to the proof of Lemma 2.2.3. For any $X \in U(N)$ with eigenvalues $(e^{i\theta_j})_{1 \leqslant j \leqslant N}$, one has, in the case where $e^{i\theta} \neq e^{i\theta_j}$ for all $j \in \{1, \dots, N\}$, and modulo π,

$$\begin{aligned}
\mathcal{I} := \mathfrak{Im}(\log(Z_X(\theta)) &= \sum_{1 \leqslant j \leqslant N} \mathfrak{Im}(\log(1 - e^{i(\theta_j - \theta)})) \\
&= \frac{1}{2} \sum_{1 \leqslant j \leqslant N} (\theta_j - \theta) + \sum_{1 \leqslant j \leqslant N} \mathfrak{Im}(\log(e^{-i(\theta_j-\theta)/2} - e^{i(\theta_j - \theta)/2})) \\
&= \frac{1}{2} \mathfrak{Im}(\log \det(X)) - \frac{N\theta}{2} + \sum_{1 \leqslant j \leqslant N} \mathfrak{Im}\left(\log\left(-2i \sin\left(\theta_j - \theta\right)/2\right)\right)) \\
&= \frac{\mathcal{J}}{2} - \frac{N(\theta + \pi)}{2}
\end{aligned}$$

where $\mathcal{J}$ denotes the version of $\mathfrak{Im}(\log \det(X))$ lying on the interval $(-\pi, \pi]$. Hence, for any $\epsilon \in (0, \pi)$, $|\mathcal{J}| \leqslant \epsilon$ if and only if $\mathcal{I}$ is on an interval of the form $\left[\frac{2k\pi - \epsilon - N(\theta+\pi)}{2}, \frac{2k\pi + \epsilon - N(\theta+\pi)}{2}\right]$ for some $k \in \mathbb{Z}$. Now, for some $A > 0$ chosen later in function of δ, let Φ be a continuous function from $\mathbb{C}$ to $[0,1]$ such that $\Phi(z) = 1$ if $|\mathfrak{Re} z - x_0| \leqslant \delta\sqrt{\log N}$ and $|\mathfrak{Im} z| \leqslant A\sqrt{\log N}$, and such that $\Phi(z) = 0$ for $|\mathfrak{Re} z - x_0| \geqslant 2\delta\sqrt{\log N}$ or $|\mathfrak{Im} z| \geqslant 2A\sqrt{\log N}$.

For $\epsilon \in (0, \pi)$, and under the Haar measure $\mathbb{P}_{U(N)}$ on $U(N)$,

$$\begin{aligned}
&\frac{\pi}{\epsilon}\mathbb{E}_{U(N)}\left[\Phi(\log(Z_X(\theta)))\mathbb{1}_{\{|\mathcal{J}|\leqslant\epsilon\}}\right] \\
&\qquad= \frac{\pi}{\epsilon}\sum_{k\in\mathbb{Z}}\mathbb{E}_{U(N)}\left[\Phi(\log(Z_X(\theta)))\mathbb{1}_{\left\{\frac{2k\pi-\epsilon-N(\theta+\pi)}{2}\leqslant\mathcal{I}\leqslant\frac{2k\pi+\epsilon-N(\theta+\pi)}{2}\right\}}\right] \\
&\qquad\qquad= \frac{\pi}{\epsilon}\sum_{k\in\mathbb{Z}}\int_{-\infty}^{\infty}dx\int_{(2k\pi-\epsilon-N(\theta+\pi))/2}^{(2k\pi+\epsilon-N(\theta+\pi))/2}dy\, D(x+iy)\Phi(x+iy) \\
&= \pi\sum_{k\in\mathbb{Z}}\int_{-\infty}^{\infty}dx\int_{-1/2}^{1/2}du\, D(x+i[k\pi-N(\theta+\pi))/2+u\epsilon])\Phi(x+i[k\pi-N(\theta+\pi))/2+u\epsilon]),
\end{aligned}$$

where D denotes the density of the law of $\log(Z_X(\theta))$, with respect to the Lebesgue measure. Now,

$$D(x+i[k\pi-N(\theta+\pi))/2+u\epsilon])\Phi(x+i[k\pi-N(\theta+\pi))/2+u\epsilon])$$

is uniformly bounded by the overall maximum of D and vanishes as soon as $|x-x_0|\geqslant 2\delta\sqrt{\log N}$ or $|k|\pi\geqslant N(|\theta|+\pi)/2+\pi/2+2A\sqrt{\log N}$. Since D and Φ are continuous functions, one can apply dominated convergence and deduce that

$$\frac{\pi}{\epsilon}\mathbb{E}_{U(N)}\left[\Phi(\log(Z_X(\theta)))\mathbb{1}_{\{|\mathcal{J}|\leqslant\epsilon\}}\right]$$

converges to

$$\pi\sum_{k\in\mathbb{Z}}\int_{-\infty}^{\infty}D(x+i[k\pi-N(\theta+\pi))/2])\Phi(x+i[k\pi-N(\theta+\pi))/2])dx$$

when ϵ goes to zero. On the other hand, if the matrix X follows $\mathbb{P}_{SU(N)}$ and if T is an independent uniform variable on $(-\pi,\pi]$, then $Xe^{iT/N}$ follows $\mathbb{P}_{U(N)}$ and its determinant is e^{iT}. One deduces:

$$\begin{aligned}
\frac{\pi}{\epsilon}\mathbb{E}_{U(N)}\left[\Phi(\log(Z_X(\theta)))\mathbb{1}_{\{|\mathcal{J}|\leqslant\epsilon\}}\right] &= \frac{\pi}{\epsilon}\mathbb{E}_{SU(N)}\left[\Phi(\log(Z_{Xe^{iT/N}}(\theta)))\mathbb{1}_{\{|T|\leqslant\epsilon\}}\right] \\
&= \frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon}\mathbb{E}_{SU(N)}\left[\Phi(\log(Z_{Xe^{it/N}}(\theta)))\right]dt \\
&= \int_{-1/2}^{1/2}\mathbb{E}_{SU(N)}\left[\Phi(\log(Z_{Xe^{2iv\epsilon/N}}(\theta)))\right]dv
\end{aligned}$$

Now, the function $X\mapsto\Phi(\log(Z_X(\theta)))$ is continuous from $U(N)$ to $[0,1]$, since Φ is continuous with compact support and $X\mapsto\log(Z_X(\theta)))$ has discontinuities only at points where its real part goes to $-\infty$. One can then apply dominated convergence and obtain:

$$\frac{\pi}{\epsilon}\mathbb{E}_{U(N)}\left[\Phi(\log(Z_X(\theta)))\mathbb{1}_{\{|\mathcal{J}|\leqslant\epsilon\}}\right]\underset{\epsilon\to 0}{\longrightarrow}\mathbb{E}_{SU(N)}\left[\Phi(\log(Z_X(\theta)))\right].$$

By comparing to the convergence obtained just above, one deduces

$$\mathbb{E}_{SU(N)}\left[\Phi(\log(Z_X(\theta)))\right]=\pi\sum_{k\in\mathbb{Z}}\int_{-\infty}^{\infty}D(x+i[k\pi-N(\theta+\pi))/2])\Phi(x+i[k\pi-N(\theta+\pi))/2])dx.$$

Since $D(z) \leqslant C_0/\log N$ and

$$\mathbb{1}_{\{|x-x_0|\leqslant\delta\sqrt{\log N},|y|\leqslant A\sqrt{\log N}\}} \leqslant \Phi(x+iy) \leqslant \mathbb{1}_{\{|x-x_0|\leqslant 2\delta\sqrt{\log N},|y|\leqslant 2A\sqrt{\log N}\}}$$

for all $x, y \in \mathbb{R}$, one deduces

$$\mathbb{P}_{SU(N)}[|\log|Z_X(\theta)| - x_0| \leqslant \delta\sqrt{\log N}, |\mathfrak{Im}\log Z_X(\theta)| \leqslant A\sqrt{\log N}] \leqslant \frac{\pi dLC_0}{\log N},$$

where $d = 4\delta\sqrt{\log N}$ is the length of the interval $[x_0 - 2\delta\sqrt{\log N}, x_0 + 2\delta\sqrt{\log N}]$ and L is the number of integers k such that $|k\pi - N(\theta+\pi))/2| \leqslant 2A\sqrt{\log N}$. Now, it is easy to check that $L \leqslant 1 + \frac{4A\sqrt{\log N}}{\pi}$, and then

$$\mathbb{P}_{SU(N)}\left[|\log|Z_X(\theta)| - x_0| \leqslant \delta\sqrt{\log N}, |\mathfrak{Im}\log Z_X(\theta)| \leqslant A\sqrt{\log N}\right] \leqslant 16\, C_0\, A\delta + \frac{4\pi\delta\, C_0}{\sqrt{\log N}}.$$

Using Lemma 2.2.2, one obtains

$$\mathbb{P}_{SU(N)}\left[|\log|Z_X(\theta)| - x_0| \leqslant \delta\sqrt{\log N}\right] \leqslant 16\, C_0\, A\delta + \frac{4\pi\delta\, C_0}{\sqrt{\log N}} + c_1' e^{-\frac{A}{2}\left(A\wedge\frac{\sqrt{\log N}}{2}\right)}.$$

Let us now choose $A := 1 + 5\log(1/\delta)$. One gets

$$A \wedge \frac{\sqrt{\log N}}{2} = [1 + 5\log(1/\delta)] \wedge \frac{\sqrt{\log N}}{2} \geqslant \frac{\sqrt{\log 2}}{2}$$

and then

$$\frac{A}{2}\left(A \wedge \frac{\sqrt{\log N}}{2}\right) \geqslant \frac{5\sqrt{\log 2}\log(1/\delta)}{4} \geqslant \log(1/\delta).$$

Therefore,

$$\mathbb{P}_{SU(N)}\left[|\log|Z_X(\theta)| - x_0| \leqslant \delta\sqrt{\log N}\right] \leqslant 16\, C_0\,\delta + 80\, C_0\,\delta\log(1/\delta) + \frac{4\pi\delta\, C_0}{\sqrt{\log N}} + c_1'\delta.$$

Since $\delta < 1/2$, one has $\delta \leqslant \delta\log(1/\delta)/\log(2)$, which implies Lemma 2.2.3, for

$$C = \frac{16\, C_0}{\log 2} + 80\, C_0 + \frac{4\pi\, C_0}{(\log 2)^{3/2}} + \frac{c_1'}{\log 2}.$$

□

2.2.5 Behaviour of the oscillation in short intervals of the log-characteristic polynomial

Lemma 2.2.7. *There exists $c_2 > 0$ such that for $\mu \in \mathbb{R}$ and $A \geqslant 0$ and uniformly in $N \geqslant M \geqslant 2 \vee \frac{|\mu|}{2\pi}$,*

$$\mathbb{P}_{SU(N)}\left(\int_0^{2\pi}\left|\mathfrak{Re}\log Z_X\left(\theta+\frac{\mu}{N}\right) - \mathfrak{Re}\log Z_X(\theta)\right|\frac{d\theta}{2\pi} \geqslant A\sqrt{\log M}\right) \leqslant \frac{c_2}{A^2},$$

$$\mathbb{P}_{SU(N)}\left(\int_0^{2\pi}\left|\mathfrak{Im}\log Z_X\left(\theta+\frac{\mu}{N}\right) - \mathfrak{Im}\log Z_X(\theta)\right|\frac{d\theta}{2\pi} \geqslant A\sqrt{\log M}\right) \leqslant \frac{c_2}{A^2}.$$

Proof. By symmetry of the problem, we can assume $\mu > 0$. Setting

$$R_\theta := \mathfrak{Re}\log Z_X\left(\theta + \frac{\mu}{N}\right) - \mathfrak{Re}\log Z_X(\theta)$$

for fixed μ (or the same expression with the imaginary part), we get:

$$\begin{aligned}
\mathbb{P}_{SU(N)}\left(\int_0^{2\pi}|R_\theta|\frac{d\theta}{2\pi} \geqslant A\sqrt{\log M}\right) &\leqslant \frac{1}{A^2\log M}\mathbb{E}_{SU(N)}\left(\left(\int_0^{2\pi}|R_\theta|\frac{d\theta}{2\pi}\right)^2\right)\\
&\leqslant \frac{1}{A^2\log M}\mathbb{E}_{SU(N)}\left(\int_0^{2\pi}R_\theta^2\frac{d\theta}{2\pi}\right)\\
&= \frac{1}{A^2\log M}\int_0^{2\pi}\mathbb{E}_{SU(N)}\left(R_\theta^2\right)\frac{d\theta}{2\pi}\\
&= \frac{1}{A^2\log M}\mathbb{E}_{U(N)}\left(R_0^2\right)\ \text{(by (2.2))}
\end{aligned}$$

Now, under $U(N)$, the canonical matrix X is almost surely unitary: let $\theta_1,\dots,\theta_N$ be its eigenangles in $[0,2\pi)$. For $j \in \{1,\dots,N\}$ and $t \in [0,2\pi)\backslash\{\theta_j\}$, we can expand the logarithm:

$$\log(1-e^{i(\theta_j-t)}) = -\sum_{k\geqslant 1}\frac{e^{ik(\theta_j-t)}}{k}$$

as a semi-convergent series. Hence, for t such that $Z_X(t) \neq 0$,

$$\log Z_X(t) = -\sum_{j=1}^{N}\sum_{k\geqslant 1}\frac{e^{ik(\theta_j-t)}}{k} = -\sum_{k\geqslant 1}\frac{e^{-ikt}}{k}\operatorname{tr}\left(X^k\right).$$

Thus:

$$\begin{aligned}
\mathfrak{Re}\log Z_X(t) &= -\frac{1}{2}\left(\sum_{k\geqslant 1}\frac{1}{k}e^{-ikt}\operatorname{tr}\left(X^k\right) + \sum_{k\geqslant 1}\frac{1}{k}e^{ikt}\operatorname{tr}\left(X^{-k}\right)\right) = -\frac{1}{2}\sum_{k\in\mathbb{Z}^*}\frac{1}{|k|}e^{-ikt}\operatorname{tr}\left(X^k\right)\\
\mathfrak{Im}\log Z_X(t) &= -\frac{1}{2i}\left(\sum_{k\geqslant 1}\frac{1}{k}e^{-ikt}\operatorname{tr}\left(X^k\right) - \sum_{k\geqslant 1}\frac{1}{k}e^{ikt}\operatorname{tr}\left(X^{-k}\right)\right) = -\frac{1}{2i}\sum_{k\in\mathbb{Z}^*}\frac{1}{k}e^{-ikt}\operatorname{tr}\left(X^k\right).
\end{aligned}$$

Here, the series in $k \in \mathbb{Z}^*$ are semi-convergent: more precisely, setting for $K \geqslant 1$,

$$S_t^{(K)} := -\frac{1}{2}\sum_{k\in\mathbb{Z}^*, |k|\leqslant K}\frac{1}{|k|}e^{-ikt}\operatorname{tr}\left(X^k\right),$$

and

$$S_t := \mathfrak{Re}\log Z_X(t),$$

$S_t^{(K)}$ tends almost surely to S_t when K goes to infinity.

Moreover, one has the following classical result ([33]): for all $p,q \in \mathbb{Z}$,

$$\mathbb{E}_{U(N)}\left(\operatorname{tr}(X^p)\overline{\operatorname{tr}(X^q)}\right) = \mathbb{1}_{\{p=q\}}\,|p| \wedge N. \tag{2.11}$$

Hence, for $K, L \geqslant 1$, $t, u \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}_{U(N)}\left(S_t^{(K)} S_u^{(L)}\right) &= \mathbb{E}_{U(N)}\left(\frac{1}{4}\sum_{p,q\in\mathbb{Z}^*,|p|\leqslant K,|q|\leqslant L}\frac{e^{-i(pt+qu)}}{|pq|}\operatorname{tr}(X^p)\operatorname{tr}(X^q)\right)\\
&= \frac{1}{4}\sum_{p,q\in\mathbb{Z}^*,|p|\leqslant K,|q|\leqslant L}\frac{e^{-i(pt+qu)}}{|pq|}\mathbb{E}_{U(N)}\left(\operatorname{tr}(X^p)\operatorname{tr}(X^q)\right)\\
&= \frac{1}{4}\sum_{p,q\in\mathbb{Z}^*,|p|\leqslant K,|q|\leqslant L}\frac{e^{-i(pt+qu)}}{|pq|}\mathbb{1}_{\{p=-q\}}\,|q|\wedge N \text{ (from (2.11))}\\
&= \frac{1}{4}\sum_{k\in\mathbb{Z}^*,|k|\leqslant K\wedge L}\frac{e^{ik(u-t)}}{k^2}\,|k|\wedge N\\
&= \frac{1}{2}\sum_{1\leqslant k\leqslant K\wedge L}\frac{k\wedge N}{k^2}\left(\frac{e^{ik(u-t)}+e^{-ik(u-t)}}{2}\right)\\
&= \frac{1}{2}\sum_{1\leqslant k\leqslant K\wedge L}\frac{k\wedge N}{k^2}\cos\left(k(u-t)\right).
\end{aligned}$$

One deduces that

$$\begin{aligned}
\mathbb{E}_{U(N)}\left((S_t^{(K)}-S_t^{(L)})^2\right) &= \mathbb{E}_{U(N)}\left((S_t^{(K)})^2\right)+\mathbb{E}_{U(N)}\left((S_t^{(L)})^2\right)-2\,\mathbb{E}_{U(N)}\left(S_t^{(K)}S_t^{(L)}\right)\\
&= \frac{1}{2}\sum_{k\geqslant 1}\frac{k\wedge N}{k^2}\cos\left(k(u-t)\right)\left(\mathbb{1}_{\{k\leqslant K\}}+\mathbb{1}_{\{k\leqslant L\}}-2\mathbb{1}_{\{k\leqslant K\wedge L\}}\right)\\
&= \frac{1}{2}\sum_{k\geqslant 1}\frac{k\wedge N}{k^2}\cos\left(k(u-t)\right)\,\mathbb{1}_{\{K\wedge L<k\leqslant K\vee L\}}\\
&\leqslant \frac{1}{2}\sum_{k\geqslant K\wedge L}\frac{k\wedge N}{k^2},
\end{aligned}$$

which tends to zero when $K\wedge L$ goes to infinity. Hence, $S_t^{(K)}$ converges in L^2 when K goes to infinity, and the limit is necessarily S_t. Therefore,

$$\mathbb{E}_{U(N)}\left(S_t S_u\right) = \lim_{K\to\infty}\frac{1}{2}\sum_{1\leqslant k\leqslant K\wedge L}\frac{k\wedge N}{k^2}\cos\left(k(u-t)\right) = \frac{1}{2}\sum_{k\geqslant 1}\frac{k\wedge N}{k^2}\cos\left(k(u-t)\right).$$

The same computation with $\widetilde{S}_t := \mathfrak{Im}\log Z_X(t)$ gives exactly the same equality:

$$\mathbb{E}_{U(N)}\left(\widetilde{S}_t\widetilde{S}_u\right) = \frac{1}{2}\sum_{k\geqslant 1}\frac{k\wedge N}{k^2}\cos\left(k(u-t)\right)$$

It is therefore enough to achieve the computations only with S_t. Using this last formula, we can write, with $\alpha = \frac{\mu}{N}$:

$$\begin{aligned}
\mathbb{E}_{U(N)}\left(R_0^2\right) &= \mathbb{E}_{U(N)}\left((S_\alpha - S_0)^2\right) = 2\mathbb{E}_U\left(S_0^2 - S_\alpha S_0\right)\\
&= \sum_{k\geqslant 1}\frac{k\wedge N}{k^2}\left(1-\cos\left(k\alpha\right)\right)
\end{aligned}$$

We can then develop $\mathbb{E}_{U(N)}\left(R_0^2\right)$:

$$\mathbb{E}_{U(N)}\left(R_0^2\right)=\sum_{k\geqslant 1}\frac{k\wedge N}{k^2}-\sum_{k\geqslant 1}\frac{k\wedge N}{k^2}\cos\left(\frac{k\mu}{N}\right)$$

But we also have

$$\sum_{k\geqslant 1}\frac{k\wedge N}{k^2}=\sum_{k=1}^{N}\frac{1}{k}+N\sum_{k>N}\frac{1}{k^2}=\log N+\gamma+O(1/N)+N\left(\frac{1}{N}+O\left(\frac{1}{N^2}\right)\right)$$

Moreover we have the following result ([48], p.37), uniformly on $\theta\in[-\pi,\pi]$:

$$\sum_{k\geqslant 1}\frac{k\wedge N}{k^2}\cos(k\theta)=-\log\left|2\sin\left(\frac{\theta}{2}\right)\right|+\mathrm{Ci}\left(N\left|\theta\right|\right)+\cos\left(N\theta\right)-\frac{\pi}{2}N\left|\theta\right|+N\theta\,\mathrm{Si}\left(N\theta\right)+O\left(\frac{1}{N}\right)\qquad(2.12)$$

where:

$$\begin{aligned}\mathrm{Si}\left(z\right)&:=\int_0^z\frac{\sin x}{x}dx=\frac{\pi}{2}-\frac{\cos z}{z}+\int_z^{+\infty}\frac{\cos x}{x^2}dx\\ \mathrm{Ci}\left(z\right)&:=-\int_z^{+\infty}\frac{\cos x}{x}dx=\gamma-\log z+\int_0^z\frac{\cos x-1}{x}dx\end{aligned}$$

Recall also that

$$\int_0^{+\infty}\frac{\sin x}{x}dx=\frac{\pi}{2}.$$

Let us denote $f\left(\mu\right):=\log\mu+\frac{\pi}{2}\mu-\cos\mu-\mathrm{Ci}\left(\mu\right)-\mu\mathrm{Si}\left(\mu\right)$. We have, for N going to infinity:

$$\begin{aligned}\mathbb{E}_{U(N)}\left(R_0^2\right)&=\log N+1+\gamma+f(\mu)-\log\mu+\log\left|2\sin\left(\frac{\theta}{2}\right)\right|+O(1/N)\quad\text{(with (2.12))}\\ &=\log N+1+\gamma+f(\mu)-\log\mu+\log\left(2\frac{\mu}{2N}\left(1+O\left(\left(\frac{\mu}{N}\right)^2\right)\right)\right)+O(1/N)\\ &=1+\gamma+f(\mu)+O\left(\left(\frac{\mu}{N}\right)^2\right)+O(1/N).\end{aligned}$$

Let us now study the behavior of the function f.

$$\begin{aligned}f(\mu)&=\log\mu-\cos\mu+\mu\left(\frac{\pi}{2}-\mathrm{Si}\left(\mu\right)\right)-\mathrm{Ci}\left(\mu\right)\\ &=\log\mu-\cos\mu+\mu\left(\frac{\cos\mu}{\mu}-\int_\mu^{+\infty}\frac{\cos x}{x^2}dx\right)-\left(\gamma-\log\mu+\int_0^\mu\frac{\cos x-1}{x}dx\right)\\ &=-\gamma-\mu\int_\mu^{+\infty}\frac{\cos x}{x^2}dx+\int_0^\mu\frac{\cos x-1}{x}dx\end{aligned}$$

One has:

$$
\begin{aligned}
f(\mu) &= -\gamma - \mu O\left(\int_{\mu}^{+\infty} \frac{1}{x^2}dx\right) + O\left(\int_0^{\mu}\left(\frac{1}{x}\wedge 1\right)dx\right) \\
&= -\gamma + O(1) + O(1+\log(\mu\vee 1)) = O\left(\log\left(\frac{\mu}{2\pi}\vee 2\right)\right),
\end{aligned}
$$

which implies

$$
\mathbb{E}_{U(N)}\left(R_0^2\right) = O\left(\log\left(\frac{\mu}{2\pi}\vee 2\right)\right) = O\left(\log M\right). \tag{2.13}
$$

□

2.2.6 Control in probability of the mean oscillation of the log-characteristic polynomials

Lemma 2.2.8. *For a certain $n\in\mathbb{N}$, let us consider an i.i.d. sequence $(U_j)_{1\leqslant j\leqslant n}$ of random matrices following the Haar measure on $SU(N)$. Let us set:*

$$
L_j(\theta) := \frac{\log\left|Z_{U_j}(\theta)\right|}{\sqrt{\frac{1}{2}\log N}}
$$

For $\delta\in(0,1/2)$, let us consider the random set:

$$
\mathscr{E}_\delta := \bigcup_{i=1}^{n}\left(\left\{\theta\in[0,2\pi]\ /\ |L_i(\theta)|\geqslant \delta^{-1}\right\}\cup\bigcup_{j\neq i}\left\{\theta\in[0,2\pi]\ /\ |L_j(\theta)-L_i(\theta)|\leqslant\delta\right\}\right)
$$

Then, there exists $c_3>0$, depending only on n, such that for all $N\geqslant 4$:

$$
\mathbb{E}\left(\lambda_{2\pi}(\mathscr{E}_\delta)\right)\leqslant c_3\delta\log(1/\delta),
$$

where $\lambda_{2\pi}$ denotes the normalised Lebesgue measure on $[0,2\pi]$.

Proof. Using Markov's inequality, one gets:

$$
\begin{aligned}
\mathbb{E}_{SU(N)}\left(\lambda_{2\pi}(\mathscr{E}_\delta)\right) &= \int_0^{2\pi}\frac{d\theta}{2\pi}\mathbb{P}_{SU(N)}\left(\theta\in\mathscr{E}_\delta\right) \\
&\leqslant \sum_{i=1}^{n}\int_0^{2\pi}\frac{d\theta}{2\pi}\mathbb{P}_{SU(N)}\left(|L_i(\theta)|\geqslant\delta^{-1}\right) \\
&\quad + \sum_{1\leqslant i\neq j\leqslant n}\int_0^{2\pi}\frac{d\theta}{2\pi}\mathbb{P}_{SU(N)}\left(|L_j(\theta)-L_i(\theta)|\leqslant\delta\right) \\
&\leqslant n\int_0^{2\pi}\frac{d\theta}{2\pi}\mathbb{P}_{SU(N)}\left(|\log|Z_X(\theta)||\geqslant\delta^{-1}\sqrt{\frac{1}{2}\log N}\right) \\
&\quad + n(n-1)\sup_{\theta\in[0,2\pi],x\in\mathbb{R}}\mathbb{P}_{SU(N)}\left(|\log|Z_X(\theta)|-x|\leqslant\delta\sqrt{\frac{1}{2}\log N}\right) \\
&\leqslant nc_1 e^{-\frac{\delta^{-1}}{\sqrt{2}}\left(\frac{\delta^{-1}}{\sqrt{2}}\wedge\frac{\sqrt{\log N}}{2}\right)} + n(n-1)C(\delta/\sqrt{2})\log(\sqrt{2}/\delta)
\end{aligned}
$$

Now,

$$e^{-\frac{\delta^{-1}}{\sqrt{2}}\left(\frac{\delta^{-1}}{\sqrt{2}}\wedge\frac{\sqrt{\log N}}{2}\right)} \leqslant e^{-\frac{\delta^{-1}}{\sqrt{2}}\left(\frac{2}{\sqrt{2}}\wedge\frac{\sqrt{\log 2}}{2}\right)} \leqslant e^{-\frac{\delta^{-1}}{5}} = O(\delta\log(1/\delta))$$

and

$$\delta\log(\sqrt{2}/\delta) \leqslant \delta\log(\sqrt{\delta^{-1}}/\delta) = \frac{3\delta}{2}\log(1/\delta),$$

which gives Lemma 2.2.8. □

2.2.7 Control in expectation of the oscillation of the log-characteristic polynomials on a small period

In the sequel, we consider the dimension $N \geqslant 4$, an integer K such that $2 \leqslant K \leqslant N/2$, defined as a function of N which is equivalent to $N/(\log N)^{3/64}$ when N goes to infinity. We denote

$$M := N/K \geqslant 2,$$

which is equivalent to $(\log N)^{3/64}$, and we also define a parameter $\delta \in (0, 1/4)$ as a function of N, equivalent to $(\log N)^{-3/32}$ when N goes to infinity. For $\theta_0 \in [0, 2\pi]$, we denote, for $0 \leqslant k \leqslant K$.

$$\theta_k := \theta_0 + \frac{2\pi k}{K} = \theta_0 + \frac{2\pi k M}{N},$$

and for $0 \leqslant k \leqslant K-1$,

$$\Delta := \theta_{k+1} - \theta_k = \frac{2\pi}{K} = \frac{2\pi M}{N}.$$

The angle θ_0 is chosen in such a way that the following technical condition is satisfied:

$$\sum_{k=0}^{K-1}\mathbb{E}_{SU(N)}\left(\left|\mathfrak{Im}\log Z_X(\theta_k + (1-\sqrt{\delta})\Delta) - \mathfrak{Im}\log Z_X(\theta_k + \sqrt{\delta}\Delta)\right|\right)$$
$$\leqslant K\mathbb{E}_{SU(N)}\left(\int_0^{2\pi}\frac{d\theta}{2\pi}\left|\mathfrak{Im}\log Z_X(\theta + (1-\sqrt{\delta})\Delta) - \mathfrak{Im}\log Z_X(\theta + \sqrt{\delta}\Delta)\right|\right).$$

This choice is always possible: indeed, if the converse (strict) inequality were true for all θ_0, then one would get a contradiction by integrating with respect to $\theta_0 \in [0, 2\pi/K)$. We then define the interval $J := [\theta_0, \theta_0 + 2\pi) = [\theta_0, \theta_K)$. Note that all the objects introduced here can be defined only as a function of N. Moreover, by applying Lemma 2.2.7 to $\theta + \sqrt{\delta}\Delta$ and $\mu = N(1 - 2\sqrt{\delta})\Delta \leqslant 2\pi M$, we deduce that the assumption made on θ_0 implies:

$$\sum_{k=0}^{K-1}\mathbb{E}_{SU(N)}\left(\left|\mathfrak{Im}\log Z_X(\theta_k + (1-\sqrt{\delta})\Delta) - \mathfrak{Im}\log Z_X(\theta_k + \sqrt{\delta}\Delta)\right|\right) = O\left(K\sqrt{\log M}\right) \tag{2.14}$$

We can then introduce the **2-oscillation** of the real and imaginary parts of the log-characteristic polynomial:

Definition 2.2.9. For $\theta \in J$ and $\mu \in [0, 2\pi M]$, and for the canonical matrix $X \in U(N)$, the 2-oscillations of $\mathfrak{Re} \log Z_X$ and $\mathfrak{Im} \log Z_X$ are defined by

$$\begin{aligned}\Delta_\mu R_\theta &:= \frac{1}{\sqrt{\log(M)}}\left|\mathfrak{Re} \log Z_X\left(\theta + \frac{\mu}{N}\right) - \mathfrak{Re} \log Z_X(\theta)\right| \\ \Delta_\mu I_\theta &:= \frac{1}{\sqrt{\log(M)}}\left|\mathfrak{Im} \log Z_X\left(\theta + \frac{\mu}{N}\right) - \mathfrak{Im} \log Z_X(\theta)\right|\end{aligned}$$

In case of several matrices $(X_j)_{1 \leqslant j \leqslant n}$, we denote the corresponding 2-oscillations by $\Delta_\mu R_\theta^{(j)}$ and $\Delta_\mu I_\theta^{(j)}$.

In the sequel, we need to introduce several random sets. The most important ones can be informally described as follows:

1. A set $\mathscr{N}_1$ of indices k such that the average of the 2-oscillations $\Delta_\mu R_\theta$ and $\Delta_\mu I_\theta$ of the log characteristic polynomials for $\theta \in [\theta_k, \theta_{k+1}]$ and $\mu \in [0, 2\pi M]$ is sufficiently small.
2. For $k \in \mathscr{N}_1$, a subset $\mathscr{G}_k$ of $[\theta_k, \theta_{k+1}]$ for which the average of the 2-oscillations with respect to $\mu \in [0, 2\pi M]$ is small enough.
3. A subset $\mathscr{N}_2$ of $\mathscr{N}_1$ of "good" indices, such that there exists $\theta \in [\theta_k, \theta_k + \sqrt{\delta}\Delta]$, both in $\mathscr{G}_k$ and $\mathscr{E}_\delta^c$. This last set, introduced in Lemma 2.2.8, corresponds to the fact that the logarithms of the absolute values of the characterize polynomials are not too large and not too close from each other: from this last condition, we can define the "carrier wave".
4. For $k \in \mathscr{N}_2$, and for some $\theta_k^* \in [\theta_k, \theta_k + \sqrt{\delta}\Delta] \cap \mathscr{G}_k \cap \mathscr{E}_\delta^c$, a subset $\mathscr{Y}_k$ of $[0, 2\pi M]$ such that the 2-oscillations $\Delta_\mu R_{\theta_k^*}$ and $\Delta_\mu I_{\theta_k^*}$ are sufficiently small. This condition will ensure that the carrier wave index corresponding to $\theta = \theta_k^* + \mu/N$ does not depend on $\mu \in \mathscr{Y}_k$.
5. From this property, we deduce that, for each pair of consecutive gaps between zeros of the carrier wave, which are sufficiently large to contain an angle of the form $\theta_k^* + \mu/N$ for $k \in \mathscr{N}_2$ and $\mu \in \mathscr{Y}_k$ ("roomy gaps"), one can find, with the notation of the introduction, a sign change of $i^N e^{iN\theta/2} F_N(e^{-i\theta})$, and then a zero of F_N.

All these sets will be precisely defined in the sequel of the paper, in a way such that their measure is "large" with "high" probability. The corresponding estimates will then be used to prove our main result.

Lemma 2.2.10. *Let $\mathbb{P}_{SU(N)}^{(n)}$ be the n-fold product of the Haar measure on $SU(N)$, $\mathbb{E}_{SU(N)}^{(n)}$ the corresponding expectation, and $(X_j)_{1 \leqslant j \leqslant n}$ the canonical sequence of n matrices in $SU(N)$. Then:*

1. *There exists a random set $\mathscr{N}_1 \subset [\![0, K-1]\!]$ such that $\mathbb{E}_{SU(N)}^{(n)}(|\mathscr{N}_1|) \geqslant (1-\delta)K$ and $\mathbb{P}_{SU(N)}^{(n)}$-a.s., $\forall\, (j,k) \in [\![1, n]\!] \times \mathscr{N}_1$,*

$$\int_{\theta_k}^{\theta_{k+1}} \int_0^{2\pi M} \Delta_\mu R_\theta^{(j)} \frac{d\mu}{2\pi M} \frac{d\theta}{2\pi} = O\left(\frac{1}{\delta K}\right)$$

and

$$\int_{\theta_k}^{\theta_{k+1}} \int_0^{2\pi M} \Delta_\mu I_\theta^{(j)} \frac{d\mu}{2\pi M} \frac{d\theta}{2\pi} = O\left(\frac{1}{\delta K}\right)$$

2. $\mathbb{P}^{(n)}_{SU(N)}$*-a.s.,* $\forall k \in \mathscr{N}_1$, $\exists \mathscr{G}_k \subset [\theta_k, \theta_{k+1})$ *such that* $\lambda_{2\pi}(\mathscr{G}_k) \geqslant (1-\delta)/K$ *and,* $\forall \theta \in \mathscr{G}_k, j \in [\![1,n]\!]$,

$$\int_0^{2\pi M} \Delta_\mu R_\theta^{(j)} \frac{d\mu}{2\pi M} = O\left(\frac{1}{\delta^2}\right) \text{ and } \int_0^{2\pi M} \Delta_\mu I_\theta^{(j)} \frac{d\mu}{2\pi M} = O\left(\frac{1}{\delta^2}\right) \tag{2.15}$$

Here, the implied constant in the $O(\cdot)$ *symbols depends only on* n.

Proof. By (2.13) and the similar estimate for the imaginary part, we have uniformly (with a universal implied constant),

$$\mathbb{E}_{U(N)}\left((\Delta_\mu R_0)^2\right) + \mathbb{E}_{U(N)}\left((\Delta_\mu I_0)^2\right) = O(1).$$

The Cauchy-Schwarz inequality ensures that

$$\mathbb{E}_{U(N)}(\Delta_\mu R_0) + \mathbb{E}_{U(N)}(\Delta_\mu I_0) = O(1),$$

i.e.

$$\int_J \mathbb{E}_{SU(N)}(\Delta_\mu R_\theta + \Delta_\mu I_\theta) \frac{d\theta}{2\pi} = O(1),$$

which implies

$$\int_J \int_0^{2\pi M} \mathbb{E}_{SU(N)}(\Delta_\mu R_\theta + \Delta_\mu I_\theta) \frac{d\mu}{2\pi M}\frac{d\theta}{2\pi} = O(1).$$

Splitting the interval J into K equal pieces and applying this estimate to n independent matrices $(X_j)_{1\leqslant j\leqslant n}$ following the Haar measure on $SU(N)$, one gets

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k=0}^{K-1} \mathbb{E}^{(n)}_{SU(N)}\left(\int_{\theta_k}^{\theta_{k+1}} \int_0^{2\pi M} (\Delta_\mu R_\theta^{(j)} + \Delta_\mu I_\theta^{(j)}) \frac{d\mu}{2\pi M}\frac{d\theta}{2\pi}\right) = O(1). \tag{2.16}$$

Applying Markov inequality, we deduce that there exists a universal constant $\kappa > 0$, such that

$$\mathbb{E}^{(n)}_{SU(N)}\left(\text{card}\left\{(j,k) \in [\![1,n]\!] \times [\![0,K-1]\!] \Big/ \int_{\theta_k}^{\theta_{k+1}} \int_0^{2\pi M} \Delta_\mu R_\theta^{(j)} \frac{d\mu}{2\pi M}\frac{d\theta}{2\pi} \geqslant \frac{\kappa n}{K\delta}\right\}\right) \leqslant \frac{\delta K}{2}$$

and

$$\mathbb{E}^{(n)}_{SU(N)}\left(\text{card}\left\{(j,k) \in [\![1,n]\!] \times [\![0,K-1]\!] \Big/ \int_{\theta_k}^{\theta_{k+1}} \int_0^{2\pi M} \Delta_\mu I_\theta^{(j)} \frac{d\mu}{2\pi M}\frac{d\theta}{2\pi} \geqslant \frac{\kappa n}{K\delta}\right\}\right) \leqslant \frac{\delta K}{2}.$$

We thus set

$$\begin{aligned}\mathscr{N}_1 := \bigcap_{j=1}^{n} \Bigg\{ k \in [\![0,K-1]\!] \Big/ \quad & \int_{\theta_k}^{\theta_{k+1}} \int_0^{2\pi M} \Delta_\mu R_\theta^{(j)} \frac{d\mu}{2\pi M}\frac{d\theta}{2\pi} \leqslant \frac{\kappa n}{K\delta}, \\ & \int_{\theta_k}^{\theta_{k+1}} \int_0^{2\pi M} \Delta_\mu I_\theta^{(j)} \frac{d\mu}{2\pi M}\frac{d\theta}{2\pi} \leqslant \frac{\kappa n}{K\delta} \Bigg\}\end{aligned}$$

and we get:

$$\mathbb{E}_{SU(N)}^{(n)}(|\mathscr{N}_1|) \geqslant (1-\delta)K. \tag{2.17}$$

Now, for $k \in \mathscr{N}_1$, let us set:

$$\mathscr{G}_k := [\theta_k, \theta_{k+1}) \cap \bigcap_{j=1}^{n} \left\{ \int_0^{2\pi M} \Delta_\mu R_\cdot^{(j)} \frac{d\mu}{2\pi M} \leqslant \frac{2\kappa n^2}{\delta^2}, \int_0^{2\pi M} \Delta_\mu I_\cdot^{(j)} \frac{d\mu}{2\pi M} \leqslant \frac{2\kappa n^2}{\delta^2} \right\}$$

Applying again Markov inequality, we get that $\mathbb{P}_{SU(N)}^{(n)}$-a.s.:

$$\lambda_{2\pi}(\mathscr{G}_k) \geqslant (1-\delta)/K. \tag{2.18}$$

$\square$

We now define good indices.

Definition 2.2.11 (Good indices)**.** An index $k \in [\![0, K-1]\!]$ is said to be **good** if :

1. $k \in \mathscr{N}_1$,
2. $\mathscr{E}_\delta^c \cap \mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta) \neq \emptyset$

We denote by $\mathscr{N}_2$ the set of good indices :

$$\mathscr{N}_2 := \left\{ k \in \mathscr{N}_1 \Big/ \mathscr{E}_\delta^c \cap \mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta) \neq \emptyset \right\} \tag{2.19}$$

An index is said to be **bad** if it is not good.

We then get the following result:

Lemma 2.2.12. *With the notation above, the set of good indices satisfies:*

$$\mathbb{E}_{SU(N)}^{(n)}(|\mathscr{N}_2|) = K\left(1 - O\left(\sqrt{\delta}\log(1/\delta)\right)\right),$$

where the implied constant in the $O(\cdot)$ symbol depends only on n.

Proof. If $k \in \mathscr{N}_2^c$, either $k \in \mathscr{N}_1^c$, or $k \in \mathscr{N}_1$ and $\mathscr{E}_\delta^c \cap \mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta) = \emptyset$, i.e. $\mathscr{N}_2^c = \mathscr{N}_1^c \cup \widetilde{\mathscr{N}_1}$ where:

$$\widetilde{\mathscr{N}_1} := \left\{ k \in \mathscr{N}_1 \Big/ \mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta) \subset \mathscr{E}_\delta \right\}.$$

By (2.17), we have $\mathbb{E}_{SU(N)}^{(n)}(|\mathscr{N}_1^c|) \leqslant \delta K$.

For all $k \in \widetilde{\mathscr{N}_1}$, we have $\mathscr{E}_\delta \supset \mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta)$, i.e. $\mathscr{E}_\delta \supset \bigcup_{k \in \widetilde{\mathscr{N}_1}} \mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta)$, where the union is disjoint, and thus, $\lambda_{2\pi}(\mathscr{E}_\delta) \geqslant \left|\widetilde{\mathscr{N}_1}\right| \min_k \lambda_{2\pi}\left(\mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta)\right)$.

$\mathbb{P}_{SU(N)}^{(n)}$-a.s., we have:

$$\begin{aligned}\lambda_{2\pi}\left(\mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta)\right) &\geqslant \lambda_{2\pi}\left(\mathscr{G}_k\right) + \lambda_{2\pi}\left([\theta_k, \theta_k + \sqrt{\delta}\Delta)\right) - \lambda_{2\pi}\left([\theta_k, \theta_k + \Delta)\right)\\ &\geqslant \frac{1}{K}\left((1-\delta) + \sqrt{\delta} - 1\right) \quad \text{by (2.18)}\end{aligned}$$

Now, since $\delta < 1/4$, we obtain $\mathbb{P}^{(n)}_{SU(N)}$-a.s.:

$$\lambda_{2\pi}\left(\mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta)\right) \geqslant \frac{\sqrt{\delta}}{2K} \tag{2.20}$$

This implies that $\mathbb{P}^{(n)}_{SU(N)}$-a.s.:

$$\left|\widetilde{\mathscr{N}_1}\right| \leqslant \frac{2K}{\sqrt{\delta}}\lambda_{2\pi}\left(\mathscr{E}_\delta\right)$$

Now, by Lemma (2.2.8), $\mathbb{E}^{(n)}_{SU(N)}\left(\left|\widetilde{\mathscr{N}_1}\right|\right) = O\left(K\sqrt{\delta}\log(1/\delta)\right)$ and then:

$$\mathbb{E}^{(n)}_{SU(N)}(|\mathscr{N}_2{}^c|) \leqslant \mathbb{E}^{(n)}_{SU(N)}(|\mathscr{N}_1{}^c|) + \mathbb{E}^{(n)}_{SU(N)}\left(\left|\widetilde{\mathscr{N}_1}\right|\right) \leqslant \delta K + O\left(K\sqrt{\delta}\log(1/\delta)\right).$$

□

2.2.8 Speed of the good oscillation of the log-characteristic polynomials

Lemma 2.2.13. *With the notation above, and $\mathbb{P}^{(n)}_{SU(N)}$-a.s., $\forall\, k \in \mathscr{N}_2$, there exists a random set $\mathscr{Y}_k \subset [0, 2\pi M]$, and $\theta_k^* \in \mathscr{E}_\delta^c \cap \mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta)$, such that*

$$\lambda_{2\pi M}\left(\mathscr{Y}_k\right) = 1 - O\left(\delta^{-2}(\log N)^{-1/4}(\log M)^{1/2}\right), \tag{2.21}$$

where λ_M is $1/2\pi M$ times the Lebesgue measure, and for all $j \in [\![1, n]\!]$, $\mu \in \mathscr{Y}_k$,

$$\Delta_\mu R^{(j)}_{\theta_k^*} = O\left(\frac{(\log N)^{1/4}}{(\log M)^{1/2}}\right) \text{ and } \Delta_\mu I^{(j)}_{\theta_k^*} = O\left(\frac{(\log N)^{1/4}}{(\log M)^{1/2}}\right) \tag{2.21}$$

Again, the implied constant in the $O(\cdot)$ symbol depends only on n.

Proof. Let $k \in \mathscr{N}_2$ and $\theta_k^* \in \mathscr{E}_\delta^c \cap \mathscr{G}_k \cap [\theta_k, \theta_k + \sqrt{\delta}\Delta)$. We set:

$$\mathscr{Y}_k := \bigcap_{j=1}^{n}\left\{\Delta_. R^{(j)}_{\theta_k^*} \leqslant \varepsilon,\ \Delta_. I^{(j)}_{\theta_k^*} \leqslant \varepsilon\right\}$$

where

$$\varepsilon := \frac{(\log N)^{1/4}}{(\log M)^{1/2}}.$$

Applying Markov inequality, we get:

$$\begin{aligned}\lambda_{2\pi M}\left(\mathscr{Y}_k^c\right) &\leqslant \lambda_{2\pi M}\left(\bigcup_{j=1}^{n}\left\{\Delta_. R^{(j)}_{\theta_k^*} \geqslant \varepsilon\right\}\right) + \lambda_{2\pi M}\left(\bigcup_{j=1}^{n}\left\{\Delta_. I^{(j)}_{\theta_k^*} \geqslant \varepsilon\right\}\right)\\ &\leqslant \frac{2n}{\varepsilon}\max_{1\leqslant j\leqslant n}\left(\int_0^{2\pi M}\Delta_\mu R^{(j)}_{\theta_k^*}\frac{d\mu}{2\pi M} \vee \int_0^{2\pi M}\Delta_\mu I^{(j)}_{\theta_k^*}\frac{d\mu}{2\pi M}\right) = \frac{1}{\varepsilon}O\left(\delta^{-2}\right),\end{aligned}$$

by (2.15) which gives the announced result.

□

2.2.9 The number of sign changes

Let us go back to Theorem 2.0.16. We need to estimate the number of zeros of F_N on the unit circle, or equivalently, the number of values of $\theta \in J$ such that the following quantity vanishes:

$$i^N e^{iN\theta/2} F_N(e^{-i\theta}) = i^N e^{iN\theta/2} \sum_{j=1}^{n} b_j \Phi_{U_{N,j}}(e^{-i\theta}) = \sum_{j=1}^{n} b_j i^N e^{iN\theta/2} Z_{U_{N,j}}(\theta). \tag{2.22}$$

Using the fact that $U_{N,j} \in SU(N)$, one checks that $i^N e^{iN\theta/2} Z_{U_{N,j}}(\theta)$ is real, and then the number of zeros of F_N on the unit circle is bounded from below by the number of sign changes, when θ increases from θ_0 to $\theta_0 + 2\pi$, of the real quantity given by the right-hand side of (2.22). Now, the order of magnitude of $\log |Z_{U_{N,j}}(\theta)|$ is $\sqrt{\log N}$ and more precisely, Lemma 2.2.8 informally means that for most values of θ, the values of $\log |Z_{U_{N,j}}(\theta)|$ for $1 \leqslant j \leqslant n$ are pairwise separated by an interval of length of order $\sqrt{\log N}$. Hence, one of the terms in the sum at the right-hand side of (2.22) should dominate all the others. If j is the corresponding index, one can expect that the sign changes of (2.22) can, at least locally, be related to the corresponding sign changes of $i^N e^{iN\theta/2} Z_{U_{N,j}}(\theta)$, which are associated to the zeros of the characteristic polynomial $Z_{U_{N,j}}$. This should give a lower bound on the number of sign changes of (2.22).

This informal discussion motivates the following definition.

Definition 2.2.14. With the notation of the previous subsections, for all $k \in \mathscr{N}_2$, we define the **carrier wave index** by:

$$j_k := \operatorname{Arg} \max_j \left\{ \mathfrak{Re} \log Z_{X_j}(\theta_k^*) \right\},$$

where θ_k^* is the random angle introduced in Lemma 2.2.13. Moreover, we consider the following interval:

$$J_k := \left[\theta_k^*, \theta_k^* + (1 - \sqrt{\delta})\Delta \right]$$

As $\theta_k^* \in \mathscr{E}_\delta^c$, we have $\forall\, j \neq j_k$, $\mathfrak{Re} \log Z_{X_j}(\theta_k^*) \leqslant \mathfrak{Re} \log Z_{X_{j_k}}(\theta_k^*) - \frac{\delta}{\sqrt{2}} \sqrt{\log N}$. From (2.21), we deduce that $\forall\, j \neq j_k$, $\forall\, \mu \in \mathscr{Y}_k$:

$$\mathfrak{Re} \log Z_{X_j}\left(\theta_k^* + \frac{\mu}{N}\right) \leqslant \mathfrak{Re} \log Z_{X_{j_k}}\left(\theta_k^* + \frac{\mu}{N}\right) - \frac{\delta}{\sqrt{2}} (\log N)^{1/2} + O\left((\log N)^{1/4}\right) \tag{2.23}$$

Now, since

$$1/\delta = O((\log N)^{1/10}),$$

with a universal implied constant, we then get, for a universal $c > 0$,

$$\frac{\left|Z_{X_j}\left(\theta_k^* + \frac{\mu}{N}\right)\right|}{\left|Z_{X_{j_k}}\left(\theta_k^* + \frac{\mu}{N}\right)\right|} \leqslant \exp\left(-2c(\log N)^{0.4} + O\left((\log N)^{1/4}\right)\right) \leqslant \exp\left(-c(\log N)^{0.4}\right),$$

for N large enough, depending only on n. This implies

$$\begin{aligned}\left|\sum_{j\neq j_k} b_j Z_{X_j}\left(\theta_k^*+\frac{\mu}{N}\right)\right| &\leqslant \frac{\sum_j |b_j|}{\min_j |b_j|}\left|b_{j_k} Z_{X_{j_k}}\left(\theta_k^*+\frac{\mu}{N}\right)\right| \exp\left(-c(\log N)^{0.4}\right)\\ &\leqslant \frac{1}{2}\left|b_{j_k} Z_{X_{j_k}}\left(\theta_k^*+\frac{\mu}{N}\right)\right|\end{aligned}$$

for $N \geqslant N_0$, where N_0 depends only on $n, b_1, \dots, b_n$.

Hence, for $k \in \mathscr{N}_2$, $\mu \in \mathscr{Y}_k$ and $\theta = \theta_k^* + \mu/N$, the quantity

$$G(\theta) := \sum_{j=1}^{n} b_j i^N e^{iN\theta/2} Z_{X_j}(\theta),$$

which is $\mathbb{P}^{(n)}_{SU(N)}$-a.s. real, has the same sign as its term of index j_k.

Theorem 2.0.16 is proven if we show that the expectation of number of sign changes of $G(\theta)$ for $\theta \in J$, under $\mathbb{P}^{(n)}_{SU(N)}$, is bounded from below by $N - o(N)$. Hence, it is sufficient to get:

$$\mathbb{E}^{(n)}_{SU(N)}\left(\sum_{k\in\mathscr{N}_2} \mathcal{S}_k\right) \geqslant N - o(N),$$

where $\mathcal{S}_k$ is the number of sign changes of $b_{j_k} i^N e^{iN\theta/2} Z_{X_{j_k}}(\theta)$, for $\theta \in J_k \cap \{\theta_k^* + \frac{\mu}{N}, \mu \in \mathscr{Y}_k\}$.

Now, for $k \in \mathscr{N}_2$, let $\alpha_{k,1} \leqslant \alpha_{k,2} \leqslant \dots \leqslant \alpha_{k,\nu_k}$ be the eigenangles, counted with multiplicity, of X_{j_k} in the interval J_k. The sign of $b_j i^N e^{iN\theta/2} Z_{X_{j_k}}$ alternates between the different intervals $(\alpha_{k,1}, \alpha_{k,2}), (\alpha_{k,2}, \alpha_{k,3}), \dots, (\alpha_{k,\nu_k-1}, \alpha_{k,\nu_k})$. Hence, for each pair of consecutive intervals containing an angle $\theta = \theta_k^* + \frac{\mu}{N}, \mu \in \mathscr{Y}_k$, we get a contribution of at least 1 for the quantity $\mathcal{S}_k$.

Every element of J_k can be written as $\theta_k^* + \frac{\mu}{N}$, for

$$0 \leqslant \mu \leqslant (1-\sqrt{\delta})N\Delta \leqslant N\Delta = 2\pi M.$$

The Lebesgue measure of the elements of J_k for which $\mu \notin \mathscr{Y}_k$ is then bounded by

$$\frac{1}{N}\lambda(\mathscr{Y}_k^c) = \frac{2\pi M}{N}\lambda_{2\pi M}(\mathscr{Y}_k^c),$$

where λ denotes the standard Lebesgue measure. Hence, if an interval $(\alpha_{k,\nu}, \alpha_{k,\nu+1})$ has a length strictly greater than this bound, it necessarily contains some $\theta = \theta_k^* + \frac{\mu}{N}$ for which $\mu \in \mathscr{Y}_k$. For some $c' > 0$ depending only on n, this condition is implied by

$$\alpha_{k,\nu+1} - \alpha_{k,\nu} > c'\frac{M}{N}\delta^{-2}(\log N)^{-1/4}(\log M)^{1/2}.$$

We will say that $(\alpha_{k,\nu}, \alpha_{k,\nu+1})$ is a **roomy gap** if this inequality is satisfied, and a **narrow gap** if

$$\alpha_{k,\nu+1} - \alpha_{k,\nu} \leqslant c'\frac{M}{N}\delta^{-2}(\log N)^{-1/4}(\log M)^{1/2}.$$

By the previous discussion, $\mathcal{S}_k$ is at least the number of pairs of consecutive roomy gaps among the intervals $(\alpha_{k,1}, \alpha_{k,2}), (\alpha_{k,2}, \alpha_{k,3}), \dots, (\alpha_{k,\nu_k-1}, \alpha_{k,\nu_k})$. If there is no narrow gap, the number of such pairs is $(\nu_k - 2)_+ \geqslant \nu_k - 2$. Moreover, if among the intervals, we replace a roomy gap by a narrow gap, this removes at most two pairs of consecutive roomy gaps. Hence, we deduce, for all $k \in \mathscr{N}_2$, that

$$\mathcal{S}_k \geqslant \nu_k - 2 - 2\psi_k,$$

where ν_k is the number of zeros of $Z_{X_{j_k}}$ in the interval J_k and ψ_k the number of narrow gaps among these zeros. Hence, we get the lower bound:

$$\mathbb{E}^{(n)}_{SU(N)}\left(\sum_{k\in\mathscr{N}_2}\mathcal{S}_k\right) \geqslant \mathbb{E}^{(n)}_{SU(N)}\left(\sum_{k\in\mathscr{N}_2}\nu_k - 2K - 2\psi\right),$$

where ψ is the total number of narrow gaps among the zeros in $[0,2\pi)$ of all the functions $(Z_j)_{1\leqslant j\leqslant n}$.

Now, $\mathbb{P}^{(n)}_{SU(N)}$-a.s., for all $k \in \mathscr{N}_2$, we have:

$$\begin{aligned}
\nu_k &= \left|\{\theta \in \left[\theta_k^*, \theta_k^* + (1-\sqrt{\delta})\Delta\right], Z_{j_k}(\theta) = 0\}\right| \\
&\geqslant \left|\{\theta \in \left[\theta_k + \sqrt{\delta}\Delta, \theta_k + (1-\sqrt{\delta})\Delta\right], Z_{j_k}(\theta) = 0\}\right| \\
&= \frac{N(1-2\sqrt{\delta})\Delta}{2\pi} + \frac{1}{\pi}\Big(\mathfrak{Im}\log Z_{X_{j_k}}(\theta_k + (1-\sqrt{\delta})\Delta) - \mathfrak{Im}\log Z_{X_{j_k}}(\theta_k + \sqrt{\delta}\Delta)\Big) \\
&\geqslant \frac{N}{K}(1-2\sqrt{\delta}) - \frac{1}{\pi}\sum_{j=1}^{n}\Big|\Big(\mathfrak{Im}\log Z_{X_j}(\theta_k + (1-\sqrt{\delta})\Delta) - \mathfrak{Im}\log Z_{X_j}(\theta_k + \sqrt{\delta}\Delta)\Big)\Big|
\end{aligned}$$

the second equality coming from Proposition 2.1.3.

Adding this inequality for all $k \in \mathscr{N}_2$, taking the expectation and using (2.14) yields the estimates :

$$\begin{aligned}
\mathbb{E}^{(n)}_{SU(N)}\left(\sum_{k\in\mathscr{N}_2}\nu_k\right) &\geqslant \frac{N}{K}(1-2\sqrt{\delta})\mathbb{E}^{(n)}_{SU(N)}(|\mathscr{N}_2|) \\
&\quad -\sum_{j=1}^{n}\sum_{k=0}^{K-1}\mathbb{E}^{(n)}_{SU(N)}\left[\Big|\Big(\mathfrak{Im}\log Z_{X_j}(\theta_k + (1-\sqrt{\delta})\Delta) - \mathfrak{Im}\log Z_{X_j}(\theta_k + \sqrt{\delta}\Delta)\Big)\Big|\right] \\
&\geqslant \frac{N}{K}(1-2\sqrt{\delta})K(1 - O(\sqrt{\delta}\log(1/\delta)) + O(K\sqrt{\log M}) \\
&\geqslant N(1 - O(\sqrt{\delta}\log(1/\delta)) + O\left(\frac{N\sqrt{\log M}}{M}\right).
\end{aligned}$$

Moreover,

$$2K = O(N/M) \tag{2.24}$$

It remains to estimate

$$\mathbb{E}^{(n)}_{SU(N)}[2\psi] = 2n\mathbb{E}_{SU(N)}[\chi] = 2n\mathbb{E}_{U(N)}[\chi],$$

where χ denotes the number of narrow gaps between the eigenvalues of the canonical unitary matrix X. The replacement of $SU(N)$ by $U(N)$ is possible since the notion of narrow gap is invariant by rotation of the eigenvalues.

Now, the last expectation can be estimated by the following result:

Lemma 2.2.15. *For $N \geqslant 1$ and $\epsilon > 0$, let U be a uniform matrix on $U(N)$ and let χ_ε be the number of pairs of eigenvalues of U whose argument differ by at most ε/N. Then, $\mathbb{E}[\chi_\varepsilon] = O(N\varepsilon^3)$.*

Proof. For $\theta_1, \theta_2 \in \mathbb{R}$, the two-point correlation density of the eigenvalues of U at $e^{i\theta_1}$ and $e^{i\theta_2}$, with respect to the uniform probability measure on the unitary group, is given by

$$\rho(e^{i\theta_1}, e^{i\theta_2}) = N^2 \left[1 - \left(\frac{\sin[N(\theta_2-\theta_1)/2]}{N\sin[(\theta_2-\theta_1)/2]}\right)^2\right].$$

Now,

$$N|\sin[(\theta_2-\theta_1)/2]| \leqslant N|\theta_2-\theta_1|/2$$

and then

$$\left(\frac{\sin[N(\theta_2-\theta_1)/2]}{N\sin[(\theta_2-\theta_1)/2]}\right)^2 \geqslant \left(\frac{\sin x}{x}\right)^2$$

for $x = N(\theta_2-\theta_1)/2$. Now, for all $x \in \mathbb{R}$, $|\sin x| \geqslant \sin|x| \geqslant |x| - |x|^3/6$, which implies

$$\left(\frac{\sin x}{x}\right)^2 \geqslant \left(1 - \frac{x^2}{6}\right)^2 \geqslant 1 - \frac{x^2}{3}$$

and

$$\rho(e^{i\theta_1}, e^{i\theta_2}) \leqslant N^2 \left[1 - \left(\frac{\sin x}{x}\right)^2\right] \leqslant \frac{N^2x^2}{3} = \frac{N^4(\theta_2-\theta_1)^2}{6}.$$

Integrating the correlation function for $\theta_1 \in [0, 2\pi)$ and $\theta' := \theta_2 - \theta_1 \in [-\varepsilon/N, \varepsilon/N]$ gives:

$$\mathbb{E}[\chi_\varepsilon] \leqslant \int_0^{2\pi} \frac{d\theta}{2\pi} \int_{-\varepsilon/N}^{\varepsilon/N} \frac{d\theta'}{2\pi} \frac{N^4(\theta')^2}{6} \leqslant N^4 \int_{-\varepsilon/N}^{\varepsilon/N} (\theta')^2 d\theta' = O\left(N^4(\varepsilon/N)^3\right).$$

□

From this result, applied for

$$\varepsilon = c' M\delta^{-2}(\log N)^{-1/4}(\log M)^{1/2}$$

we get the estimate:

$$\mathbb{E}^{(n)}_{SU(N)}[2\psi] = O(N\varepsilon^3) = O\left(NM^3\delta^{-6}(\log N)^{-3/4}(\log M)^{3/2}\right). \tag{2.25}$$

The estimates (2.24), (2.24) and (2.25) imply:

$$\mathbb{E}^{(n)}_{SU(N)}\left(\sum_{k\in\mathcal{N}_2} \mathcal{S}_k\right) \geqslant N\left[1 - O\left(\sqrt{\delta}\log(1/\delta) + \frac{\sqrt{\log M}}{M} + M^3\delta^{-6}(\log N)^{-3/4}(\log M)^{3/2}\right)\right]$$

From the values taken for δ and M, we get:

$$\begin{aligned}\sqrt{\delta}\log(1/\delta) &= O\left((\log N)^{-3/64}\log\log N\right)\\ \frac{\sqrt{\log M}}{M} &= O\left(\sqrt{\log\log N}(\log N)^{-3/64}\right)\end{aligned}$$

and

$$\begin{aligned}M^3\delta^{-6}(\log N)^{-3/4}(\log M)^{3/2} &= O\left((\log N)^{9/64}(\log N)^{18/32}(\log N)^{-3/4}(\log\log N)^{3/2}\right)\\ &= O\left((\log N)^{-3/64}(\log\log N)^{3/2}\right)\end{aligned}$$

Finally, we get

$$\mathbb{E}^{(n)}_{SU(N)}\left(\sum_{k\in\mathscr{N}_2}\mathcal{S}_k\right) = N\left(1-O\left((\log N)^{-1/22}\right)\right)$$

which completes the proof of Theorem 2.0.16.

Chapter 3

Mod-* convergence

This chapter is devoted to the study of mod-* convergence, a new type of convergence in probability theory (see definition 3.2.1). We first remind the main notions and theorems of the theory and then construct a particular sequence of random variables that converges in the mod-* sense. Using these random variables, we will be able to metrize the convergence (see chapter 4). Here, we focus on an arithmetic application of this construction, the creation of a model made of independent random variables conditioned in a certain way that reproduces the mod-* fluctuations in a celebrated theorem of Selberg and Sathé about the number of prime divisors of a random uniform integer. We finally propose an explanation of the splitting phenomenon occuring in the mod-* limit of this latest theorem by means of an additional randomisation.

3.1 Introduction

The mod-Gaussian convergence concept arose as a refinement of the usual Central Limit Theorem (CLT). The fact that this theorem occurs in several different situations does not mean that the same reasons are hidden behind. Indeed, the most famous statement about the CLT concerns the sum of independent random variables having a second moment, i.e.

Theorem 3.1.1 (CLT). *1. Let $(X_k)_k$ be a sequence of i.i.d. random variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{E}\left(X_1^2\right) = 1$. Then*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

2. (Lindberg-Feller) If $(X_k)_k$ is a sequence of independent random variables with $\mathbb{E}(X_k) = 0$

and $\mathbb{E}\left(X_1^2\right) = \sigma_k^2$*, let* $S_n^2 := \sum_{k=1}^n \sigma_k^2 = \mathbb{E}\left(\left(\sum_{k=1}^n X_k\right)^2\right)$*. Suppose that*

$$\frac{1}{S_n^2}\sum_{k=1}^n \mathbb{E}\left(X_k^2 \mathbb{1}_{\{|X_k|\geqslant tS_n\}}\right) \xrightarrow[n\to+\infty]{} 0$$

Then

$$\frac{1}{S_n}\sum_{k=1}^n X_k \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}(0,1)$$

But the Gaussian distribution has a wide range of universality, and the independance does not play a crucial rôle, as one can see by considering the following

Theorem 3.1.2 (Salem-Zygmund ([105], Sect. XVI-5))**.** *Let* $U \sim \mathscr{U}([0,1])$ *and*

$$X_k := r_k \cos\left(2\pi n_k U + a_k\right)$$

with $0 \leqslant a_k < 2\pi$*, where* $(n_k)_k$ *satisfies the following lacunary condition*

$$\exists\,\delta > 0 \ \ / \ \ \forall\, k \geqslant 1, \ \ n_{k+1} \geqslant (1+\delta)n_k$$

and where $(r_k)_k$ *satisfies the conditions*

$$(i) \quad \sum_{k\geqslant 0} r_k^2 = +\infty$$

$$(ii) \quad \frac{r_k^2}{r_1^2+\cdots+r_k^2} \xrightarrow[k\to+\infty]{} 0$$

Then, we have the following CLT :

$$\frac{1}{\sqrt{\sum_{k=1}^n r_k^2}}\sum_{k=1}^n X_k \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}\left(0,\frac{1}{2}\right)$$

This CLT also holds for linearly independent $(n_k)_k$ (see [59] p. 47) or for more general classes of functions than the trigonometric ones : orthogonal systems of uniformly bounded functions ([76]), Lipschitz functions ([58]), Hölder functions ([40]), etc. What these theorems show is that a sum of random variables with the same exact source of randomness (whence terms cannot be more dependant) act as a sum of independent random variables after a proper renormalisation.

Indeed, this is more the absence of correlation than the absence of dependance that explains the apparition of the Gaussian distribution. Consequently, if a general CLT captures its universality, it is too general to capture the specificity of a particular problem, and the question of the characterisation of a distribution with a weaker renormalisation becomes more relevant. A fundamental example illustrates the problem (see e.g. [4] and references cited)

Theorem 3.1.3 (Law of small numbers)**.** *Let* $(B_k)_{k\geqslant 1}$ *be a sequence of independent Bernoulli random variables with* $\mathbb{P}(B_k = 1) = p_k = 1 - \mathbb{P}(B_k = 0)$*. Suppose moreover that*

$$(i) \quad \sum_{k\geqslant 1} p_k = +\infty$$

$$(ii) \quad \sum_{k\geqslant 1} p_k^2 < +\infty$$

Then, we have the following CLT :

$$\frac{1}{\sqrt{\sum_{k=1}^{n} p_k}} \left(\sum_{k=1}^{n} B_k - \sum_{k=1}^{n} p_k \right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathscr{N}(0,1)$$

and the following total variation distance with a Poisson distribution

$$d_{\mathrm{TV}} \left(\sum_{k=1}^{n} B_k, \mathscr{P}\left(\sum_{k=1}^{n} p_k \right) \right) \leqslant \frac{1}{\sqrt{\sum_{k=1}^{n} p_k}} \xrightarrow[n \to +\infty]{} 0$$

On the point of view of the total variation distance, the integer-valued random variable $\sum_{k=1}^{n} B_k$ looks like a Poisson distributed random variable more than a Gaussian, but the parameter of the Poisson distribution grows to infinity, and a further renormalisation is needed to make it converge.

This last distributional approximation in terms of total variation distance has a counterpart for the sum of i.i.d. random variables, the celebrated *Berry-Esséen theorem*. Indeed, if the CLT asserts that

$$d_{\mathrm{Kol}} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k, \mathscr{N}(0,1) \right) := \sup_{x \in \mathbb{R}} \left| \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \leqslant x \right) - \int_{-\infty}^{x} e^{-u^2/2} \frac{du}{\sqrt{2\pi}} \right| \xrightarrow[n \to +\infty]{} 0$$

the Berry-Esséen theorem can be tougth of as its direct continuation since it provides the rate of convergence of this latest limit.

Theorem 3.1.4 (Berry-Esséen bound, [17, 36])**.** *Let $(X_k)_k$ be a sequence of i.i.d. random variables with $\mathbb{E}(X_1) = 0$, $\mathbb{E}\left(X_1^2\right) = 1$ and $\mathbb{E}\left(|X_1|^3\right) < \infty$. Then*

$$d_{\mathrm{Kol}} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k, \mathscr{N}(0,1) \right) \leqslant \frac{3\,\mathbb{E}\left(|X_1|^3\right)}{\sqrt{n}}$$

This is a first possible refinement of the CLT, at the "second order". Another such refinement is the local CLT, which is in particular relevant when no second moment is available (for Cauchy distributions for instance). For example, consider a random variable Y_α selected according to the symmetric stable distribution of parameter $\alpha \in (0,2)$, i.e.

$$\mathbb{E}\left(e^{itY_\alpha}\right) = e^{-|t|^\alpha}$$

Then, if $(X_k)_k$ is a sequence of i.i.d. random variables equal in law to Y_α, the usual CLT fails (there is no first moment, as one can see by differentiating the last equality). But one can renormalise something else than $\sum_{k=1}^{n} X_k$, in particular, one can renormalise the probability of being in a given set.

Theorem 3.1.5 (Local Central Limit Theorem, [29])**.** *Let $(X_k)_k$ be a sequence of symmetric random variables that are not distributed on a lattice. Suppose that there exists $(b_k)_k$ such that $b_k \to +\infty$ when $k \to +\infty$ and that*

$$\frac{1}{b_n} \sum_{k=1}^{n} X_k \xrightarrow[n \to +\infty]{\mathcal{L}} Y_\alpha$$

Then, for all B borelian set of $\mathbb{R}$ relatively compact such that $\lambda(\partial B) = 0$ (λ being the Lebesgue measure), we have

$$b_n \,\mathbb{P}\left(\sum_{k=1}^{n} X_k \in B\right) \xrightarrow[n\to+\infty]{} s_\alpha \lambda(B) \quad \textit{with } s_\alpha := \int_{\mathbb{R}} e^{-|t|^\alpha} \frac{dt}{2\pi}$$

If moreover $(c_n)_n$ is a sequence such that $c_n \to +\infty$ and $c_n = o(b_n)$ when $n \to \infty$, then

$$\frac{b_n}{c_n} \,\mathbb{P}\left(\frac{1}{c_n}\sum_{k=1}^{n} X_k \in B\right) \xrightarrow[n\to+\infty]{} s_\alpha \lambda(B)$$

This last statement means that the sum is still in the domain of attraction of the local CLT in the regime below b_n. But if $c_n \neq o(b_n)$, we change the regime for a *large deviation regime*, and the relevant theorems are then the ones of Cramer (in logarithmic scale) and Bahadur-Rao (in the classical scale).

For $X \in \mathbb{R}$ with a non degenerate distribution such that $\mathbb{E}\left(e^{yX}\right) < \infty$ for $y \in I \subset \mathbb{R}$, we set

$$\begin{aligned} \Lambda_X(y) &:= \log \mathbb{E}\left(e^{yX}\right) \\ \Lambda_X^*(x) &:= \mathrm{Leg}(\Lambda_X)(x) := \sup_{y\in\mathbb{R}} \{xy - \Lambda_X(y)\} \\ \mathscr{D}_{\Lambda_X} &:= \{\Lambda_X < \infty\} \end{aligned}$$

We suppose moreover that there exists $c > 0$ such that $(-c, c) \subset \mathscr{D}_\Lambda$. Then, we have the following (see [39] and references cited)

Theorem 3.1.6 (Large deviations)**.** *Let $(X_k)_k$ be a sequence of i.i.d. symmetric random variables that are not distributed on a lattice such that their law satisfies $\mathbb{E}\left(e^{yX}\right) < \infty$ for $y \in I \subset \mathbb{R}$.*

1. (Cramer) *Suppose that Λ_X is $\mathcal{C}^1$ on $\mathring{\mathscr{D}}_{\Lambda_X}$. Let $\xi \in \mathring{\mathscr{D}}_{\Lambda_X}$, $\xi > 0$. Then*

$$\frac{1}{n} \log \mathbb{P}\left(\sum_{k=1}^{n} X_k \geqslant \Lambda_X'(\xi)\right) \xrightarrow[n\to+\infty]{} -\Lambda_X^*(\xi)$$

2. (Bahadur-Rao) *Suppose moreover that Λ_X is $\mathcal{C}^2$ on $\mathring{\mathscr{D}}_{\Lambda_X}$. Let $\xi \in \mathring{\mathscr{D}}_{\Lambda_X}$, $\xi > 0$. Then*

$$\xi \sqrt{2\pi n \Lambda_X''(\xi)}\; e^{n\Lambda_X^*(\xi)}\; \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k \geqslant \Lambda_X'(\xi)\right) \xrightarrow[n\to+\infty]{} 1$$

The first statement is a consequence of the second with a change of renormalisation.

The mod-* convergence is a general framework that will imply first order renormalisation limit (in distribution, i.e. CLT's) but also the different types of second order renormalisation that were encountered (local CLT, large deviations, Kolmogorov or total variation approximation). It will moreover allow to partition into *subclasses of universality* the *global* class of universality of the Gaussian distribution, i.e. a suitable change of normalisation will reveal a dependance that vanishes with a further renormalisation and that will allow to classify random variables that belong to the domain of attraction of the Gaussian according to this second-order dependance.

3.2 The mod-* convergence

3.2.1 The mod-Gaussian convergence

Definition 3.2.1 (Mod-gaussian convergence)**.** Let $(Z_n)_n$ be a sequence of random variables of expectation 0 and $(\gamma_n)_n$ be a sequence of strictly positive real numbers. Let $G \sim \mathscr{N}(0,1)$. We say that $(Z_n)_n$ converges in the mod-gaussian sense if

$$\frac{\mathbb{E}\left(e^{iuZ_n}\right)}{\mathbb{E}\left(e^{iu\gamma_n G}\right)} \xrightarrow[n\to+\infty]{} \Phi(u)$$

the convergence being locally uniform in $u \in \mathbb{R}$ and $\Phi : \mathbb{C} \to \mathbb{C}$ hence being a continuous function satisfying $\Phi(0) = 1$ and $\overline{\Phi(u)} = \Phi(-u)$ where $z \mapsto \overline{z}$ denotes the complex conjugation.

When such a convergence holds, we write it as

$$(Z_n, \gamma_n) \xrightarrow[n\to+\infty]{\text{mod-G}} \Phi$$

Remark 3.2.2. One can always be reduced to the case of a sequence of random variables with zero expectation. Otherwise, we include additional renormalization in the Fourier transform of the Gaussian random variable, which corresponds to the original definition of [53].

Remark 3.2.3. As a direct corollary of the definition, if $\gamma_n \to +\infty$ when $n \to +\infty$, then,

$$\frac{Z_n}{\sqrt{\gamma_n}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}(0,1) \quad \text{if } (Z_n, \gamma_n) \xrightarrow[n\to+\infty]{\text{mod-G}} \Phi$$

A trivial example but a useful insight for the intuition allows to illustrate the concept :

Example 3.2.4. Consider $Z_n := Y_n + G_n$ where $(Z_n)_n$ is independent of $(G_n)_n$, with $G_n \stackrel{\mathcal{L}}{=} \gamma_n G$, with $G \sim \mathscr{N}(0,1)$ and $Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} Y_\infty$, . Then,

$$\frac{\mathbb{E}\left(e^{iuZ_n}\right)}{\mathbb{E}\left(e^{iu\gamma_n G}\right)} = \mathbb{E}\left(e^{iuY_n}\right) \xrightarrow[n\to+\infty]{} \Phi(u) := \mathbb{E}\left(e^{iuY_\infty}\right)$$

Thus, in the case of an *additive* independent gaussian noise, such a renormalisation gives at the limit the Fourier transform of a probability measure.

An interesting question to ask concerning this particular type convergence concerns its possible interpretation in terms of probability. Indeed, the intuitive idea that the general case would deal with an additive correlated noise that disappears with this particular type of renormalisation is not satisfactory if we escape from the domain of probability theory at the limit ; this is still the case in the last example since the limiting function is still the Fourier transform of a probability distribution, but in the general case, the limiting function belongs to a much wider functionnal space, as one can see by looking at the example 3.4.8.

One possible solution to this problem (with additional restrictive conditions) will be given in section 3.4. Other questions of interest concerning the limiting function Φ are still to be answered, and in particular what quantity of information on the sequence $(Z_n)_n$ is contained in Φ ? For example, is it an entropy of large deviations or is there more ?

3.2.2 The mod-Poisson convergence

Recall that if $P_\gamma \sim \mathscr{P}(\gamma)$ is a Poisson distributed random variable, $\mathbb{P}(P_\gamma = k) = e^{-\gamma}\gamma^k/k!$ which is equivalent to $\mathbb{E}\left(e^{iuP_\gamma}\right) = \exp\left(\gamma(e^{iu}-1)\right)$. In the same vein as before, we define the mod-Poisson convergence by the following :

Definition 3.2.5. Let $(Z_n)_n$ be a sequence of positive random variables and $(\gamma_n)_n$ be a sequence of strictly positive real numbers. $(Z_n)_n$ is said to converge mod-Poisson with parameters $(\gamma_n)_n$ if for all $\theta \in \mathbb{R}$

$$\frac{\mathbb{E}\left(e^{i\theta Z_n}\right)}{\mathbb{E}\left(e^{i\theta P_{\gamma_n}}\right)} \xrightarrow[n\to+\infty]{} \Phi\left(e^{i\theta}\right)$$

where $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous function satisfying $\Phi(1) = 1$, the last convergence being locally uniform in $\theta \in \mathbb{R}$, and $P_{\gamma_n} \sim \mathscr{P}(\gamma_n)$.

When such a convergence holds, we write it as

$$(Z_n, \gamma_n) \xrightarrow[n\to+\infty]{\text{mod-P}} \Phi$$

Note that although the convergence holds for $\theta \in \mathbb{R}$, we require the function Φ to be positive on $\mathbb{R}_+$, which is an additional restriction of the definition given in [53].

Example 3.2.6. Let $(B_k)_k$ be a sequence of Bernoulli random variables satisfying the hypotheses of theorem 3.1.3, let $Z_n := \sum_{k=1}^n B_k$ and $\gamma_n := \sum_{k=1}^n p_k$. Then,

$$(Z_n, \gamma_n) \xrightarrow[n\to+\infty]{\text{mod-P}} \Phi$$

where

$$\Phi(x) = \prod_{k\geqslant 1}(1+p_k(x-1))e^{-p_k(x-1)} \tag{3.1}$$

Indeed, setting $P_{\gamma_n} \sim \mathscr{P}(\gamma_n)$ one has

$$\frac{\mathbb{E}\left(x^{Z_n}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} = \frac{\prod_{k=1}^n \mathbb{E}\left(x^{B_k}\right)}{e^{\gamma_n(x-1)}} = \prod_{k=1}^n \left(1+p_k(x-1)\right)e^{-p_k(x-1)} \xrightarrow[n\to+\infty]{} \prod_{k\geqslant 1}(1+p_k(x-1))e^{-p_k(x-1)}$$

since $\sum_k p_k^2 < \infty$ and $(1+p_k(x-1))e^{-p_k(x-1)} = \exp\left(-p_k^2(x-1)^2/2 + o(p_k^2)\right)$.

Remark 3.2.7. The limiting function Φ is not unique, since it is defined up to a multiplication by an exponential (see [53]). Using a product representation of the exponential in certain cases of parameters p_k, one can have a different form of the limiting function for a different factor γ_n.

The following proposition allows to understand the links between mod-Gaussian and mod-Poisson convergence (see [4]) :

Proposition 3.2.8. *Let $(Z_n)_n$ be a sequence of random variables such that $(Z_n, \gamma_n)_n \xrightarrow[n\to+\infty]{\text{mod-P}} \Phi$ for a certain sequence $(\gamma_n)_n$ of strictly positive reals. Then*

1. $\frac{Z_n}{\gamma_n} \xrightarrow[n\to+\infty]{} 1$ *in probability, i.e.*

$$\forall \varepsilon > 0, \quad \mathbb{P}\left(\left|\frac{Z_n}{\gamma_n} - 1\right| \geqslant \varepsilon\right) \xrightarrow[n\to+\infty]{} 0$$

2. We have the CLT

$$\frac{Z_n - \gamma_n}{\sqrt{\gamma_n}} \xrightarrow[n\to+\infty]{} \mathscr{N}(0,1)$$

3. If moreover $Z_n \in \mathbb{Z}$ *for all* $n \in \mathbb{N}$*, and if* $u \mapsto \mathbb{E}\left(e^{iuZ_n}\right) \in \mathcal{C}^1$ *for all* $n \in \mathbb{N}$ *and if in addition the mod-Poisson convergence occur in the* C^1 *topology, then*

$$d_{\mathrm{Kol}}\left(Z_n, \mathscr{P}(\gamma_n)\right) \leqslant \frac{\|\Phi'\|_\infty}{\sqrt{\gamma_n}}$$

Z_n is hence close to $\mathscr{P}(\gamma_n)$ in the sense of the Kolmogorov distance if a mod-Poisson convergence occurs in the sense of the C^1 topology. But $\mathscr{P}(\gamma_n)$ is a divergent sequence of distributions. One can always write the triangle inequality

$$\begin{aligned} d_{\mathrm{Kol}}\left(Z_n, \mathscr{P}(\gamma_n)\right) &= d_{\mathrm{Kol}}\left(\frac{Z_n - \gamma_n}{\sqrt{\gamma_n}}, \frac{\mathscr{P}(\gamma_n) - \gamma_n}{\sqrt{\gamma_n}}\right) \\ &\leqslant d_{\mathrm{Kol}}\left(\frac{Z_n - \gamma_n}{\sqrt{\gamma_n}}, G\right) + d_{\mathrm{Kol}}\left(\frac{\mathscr{P}(\gamma_n) - \gamma_n}{\sqrt{\gamma_n}}, G\right) \end{aligned}$$

and since $\left(\mathscr{P}(\gamma_n) - \gamma_n\right)/\sqrt{\gamma_n} \xrightarrow[n\to+\infty]{\mathcal{L}} G \sim \mathscr{N}(0,1)$, the limiting term goes to zero. But the speed of convergence involving Φ is more accurate (for a similar phenomenon in the Gaussian setting, see theorem 4.7.1).

3.2.3 Local limit theorems

Following [29], one can weaken the definitions of mod-* convergence. Let $d \geqslant 1$ and $X_n \in \mathbb{R}^d$. Let $\mu \in \wp(\mathbb{R}^d)$ (probability measure on $\mathbb{R}^d$) of characteristic function φ, i.e. $\varphi(t) = \int_{\mathbb{R}^d} e^{it^*x}\mu(dx)$ where $t^*x := \sum_{k=1}^d x_k t_k$. Define $\varphi_n(t) := \mathbb{E}\left(e^{it^*X_n}\right)$ and suppose that

(H_1) $\quad \varphi \in L^1(\lambda) \iff \mu \ll \lambda$

(H_2) $\quad \exists\, (A_n)_n \in GL_d(\mathbb{R})^{\mathbb{N}}$ s.t. $\Sigma_n := A_n^{-1} \xrightarrow[n\to\infty]{} 0$ and

$$t \mapsto \varphi_n(\Sigma_n^* t) \xrightarrow[n\to\infty]{} \varphi \text{ locally uniformly in } t \iff \Sigma_n^* X_n \xrightarrow[n\to\infty]{} \mu$$

(H_3) $\quad$ for all $k \geqslant 0$, $f_{n,k} : t \mapsto \varphi_n(\Sigma_n^* t)\mathbb{1}_{\{|\Sigma_n^* t| \leqslant k\}}$ is uniformly integrable, i.e.

$$\lim_{a\to+\infty} \sup_{n\geqslant 1} \int_{|t|\geqslant a} |f_{n,k}(t)|\, dt = 0$$

Definition 3.2.9 (Mod- φ convergence, [29])**.** If $(H1)$, $(H2)$ and $(H3)$ are satisfied, we say that there is mod- φ convergence.

Note that one can replace (H_2) by

(H_2') $\exists\,(A_n)_n \in GL_d(\mathbb{R})^{\mathbb{N}}$ s.t. $\Sigma_n := A_n^{-1} \xrightarrow[n\to\infty]{} 0$, $\exists\,\Phi \in \mathcal{C}^0(\mathbb{R}^d, \mathbb{C})$ with compact support s.t. $\Phi(0) = 1$ and s.t. for all t with $|\Sigma_n^* t| \leqslant k$

$$\varphi_n(t) = \Phi(t)\,\varphi(A_n^* t)\,(1 + o(1))$$

Such a notion of convergence is a mimic of the Gaussian and Poisson case that applies for stable distributions. Hence, one can try to find local limit theorems when no moments are available.

Theorem 3.2.10 (Local limit theorem with mod-φ converging sequences)**.** *Let $(X_n)_n$ be a sequence converging in the mod-φ sense. Then, for all $f \in \mathcal{C}^0(\mathbb{R}^d, \mathbb{C})$ compactly supported*

$$|\det(A_n)|\,\mathbb{E}\left(f(X_n)\right) \xrightarrow[n\to\infty]{} \frac{d\mu}{d\lambda}(0) \int_{\mathbb{R}^d} f\,d\lambda$$

In particular, for B a compact borelian set such that $\lambda(\partial B) = 0$,

$$|\det(A_n)|\,\mathbb{P}\left(X_n \in B\right) \xrightarrow[n\to+\infty]{} \frac{d\mu}{d\lambda}(0)\lambda(B)$$

This theorem allows to find automatically a local limit theorem for sequences converging in the mod-φ sense, in particular if φ is the Fourier transform of a Stable distribution. Several famous examples of convergence towards a Cauchy distribution arise in probability, with an explicit computation of the Fourier transform, and this is the case for the winding number of the planar Brownian motion.

Example 3.2.11. Let $W := (W_t)_{t\geqslant 0}$ be the complex (or planar) Brownian motion starting in $1 \equiv (1, 0)$. One has the polar representation

$$W_t = R_t e^{i\theta_t}$$

where $R_t := |W_t|$ is the modulus of the Brownian motion and $\theta_t = \arg(W_t)$ is the winding number of the Brownian motion.

Spitzer ([91]) computed explicitely the Fourier transform of θ_t :

$$\mathbb{E}\left(e^{ix\theta_t}\right) = \sqrt{\frac{\pi}{8t}}\, e^{-\frac{1}{4t}} \left(I_{\frac{|x|-1}{2}}\left(\frac{1}{4t}\right) + I_{\frac{|x|+1}{2}}\left(\frac{1}{4t}\right)\right)$$

where I_ν designates the Bessel function of first type, defined by

$$I_\nu(z) := \sum_{k\geqslant 0} \frac{1}{k!\,\Gamma(\nu + k + 1)} \left(\frac{z}{2}\right)^{\nu+2k}$$

He deduced the following theorem

$$\frac{2\theta_t}{\log t} \xrightarrow[t\to+\infty]{\mathcal{L}} \mathscr{C}(1)$$

where $\mathscr{C}(1)$ is the symmetric Cauchy random variable of parameter 1, with density $x \mapsto \frac{1}{\pi(1+x^2)}$ and characteristic function $\varphi(t) := e^{-|t|}$.

Since the Fourier transform is explicit, one can compute a mod-φ convergence (here mod-Cauchy) to get :

Corollary 3.2.12 (Local limit theorem for the winding number of the planar Brownian motion,[29])**.** *For all $a < b$ real numbers*

$$\frac{\log(t)}{4}\,\mathbb{P}\left(a \leqslant \theta_t \leqslant b\right) \xrightarrow[t\to+\infty]{} \frac{b-a}{2\pi}$$

Note that we do not have the uniform distribution arising at the limit : a and b are in $\mathbb{R}$ and not on the circle, i.e. the Brownian motion turns a lot around the origin and not only of a fraction of 2π.

3.2.4 Precise large deviations

This is the classical next step once escaped from the regime of local limit theorems.

Let $(X_n)_n$ be a sequence of random variables with $\varphi_n(z) := \mathbb{E}\left(e^{zX_n}\right)$, $z \in \mathbb{C}$. We suppose that φ_n is defined on an open neighbourhood of 0, i.e. on S_c for a certain $c > 0$, where

$$S_c := \{-c < \mathfrak{Re} < c\}$$

We moreover suppose that there exists a random variable Y with infinitely divisible distribution and a function $\eta : S_c \to \mathbb{C}$ such that

$$\mathbb{E}\left(e^{zY}\right) = e^{\eta(z)}$$

Last, we suppose that there exists an analytic function ψ which does not vanishes in S_c such that locally uniformly in $z \in S_c$, one has

$$e^{-t_n\eta(z)}\varphi_n(z) = \frac{\mathbb{E}\left(e^{zX_n}\right)}{\left[\mathbb{E}\left(e^{zY}\right)\right]^{t_n}} \xrightarrow[n\to+\infty]{} \psi(t) \tag{3.2}$$

with $t_n \to +\infty$ when $n \to \infty$.

Two cases are possible according to the periodicity of the Fourier transforms, i.e. whether X_n and Y have a distribution concentrated on a lattice or not.

Theorem 3.2.13 (Precise large deviations, [39])**.** *Let η^* be the Legendre transform of η. Then*

- *If the X_n's are on a lattice,*

  1. *if $x \in \eta'\left(]-c, c[\right)$ and $t_n x \in \mathbb{N}$, there exist $\nu \in \mathbb{N}^*$ and $(a_k)_{k<\nu}$ s.t.*

$$\mathbb{P}\left(X_n = t_n x\right) = e^{-t_n\eta^*(x)}\sqrt{\frac{(\eta^*)''(x)}{2\pi t_n}}\left[\psi \circ (\eta^*)''(x) + \sum_{k=1}^{\nu-1}\frac{a_k}{t_n^k} + O\left(\frac{1}{t_n^\nu}\right)\right]$$

2. *if $x \in \eta'([0,c[)$, there exist $\nu \in \mathbb{N}^*$ and $(b_k)_{k<\nu}$ s.t.*

$$\mathbb{P}(X_n \geqslant t_n x) = \frac{e^{-t_n \eta^*(x)}}{\sqrt{2\pi t_n\, \eta'' \circ (\eta')^{-1}(x)}} \left[\psi \circ (\eta^*)''(x) + \sum_{k=1}^{\nu-1} \frac{b_k}{t_n^k} + O\left(\frac{1}{t_n^\nu}\right) \right]$$

- *If the X_n's are not on a lattice and if $x \in \eta'([0,c[)$, then,*

$$\mathbb{P}(X_n \geqslant t_n x) = \frac{\psi \circ (\eta')^{-1}(x)}{(\eta')^{-1}(x)} \frac{e^{-t_n \eta^*(x)}}{\sqrt{2\pi t_n \eta'' \circ (\eta')^{-1}(x)}} (1 + o(1))$$

The number ν is the speed of convergence of (3.2).

3.3 Fundamental examples

3.3.1 Random permutations

Let $\sigma \sim \mathbb{P}_\theta^{(n)}$ where $\mathbb{P}_\theta^{(n)}$ is the Ewens measure defined in (1.4). Recall that the cycle structure is the vector $\mathcal{C}(\sigma) := (c_k(\sigma))_{1 \leqslant k \leqslant n}$ where $c_k(\sigma)$ denotes the number of k-cycles of σ and that the law of the cycle structure under the Ewens measure is explicitely given by the following equality in law

$$(c_1, \dots, c_n) \stackrel{\mathcal{L}}{=} \left(P_1, \dots, P_n \middle| \textstyle\sum_{k=1}^n k P_k = n\right), \quad P_k \sim \mathscr{P}\left(\frac{\theta}{k}\right)$$

the $(P_k)_k$ being independent. Moreover, because of the equality

$$\sum_{\sigma \in \mathfrak{S}_n} \theta^{C(\sigma)} = \theta(\theta+1)\dots(\theta+n-1)$$

one has

$$\mathbb{E}_\theta^{(n)}\left(x^C\right) = \prod_{k=0}^{n-1} \frac{x\theta + k}{\theta + k} = \prod_{k=1}^{n} \mathbb{E}\left(x^{B_k(\theta)}\right)$$

with independent Bernoulli random variables given by

$$\mathbb{P}(B_k(\theta) = 1) = \frac{\theta}{\theta + k - 1} = 1 - \mathbb{P}(B_k(\theta) = 0)$$

Since the total number of cycles is a sum of independant random variables with finite variance, one has the usual CLT, due in this case to Goncharov (1942) :

$$\frac{C - \mathbb{E}_\theta^{(n)}(C)}{\sqrt{\mathrm{Var}_\theta^{(n)}(C)}} = \frac{\sum_{k=1}^n B_k(\theta) - \theta \sum_{k=1}^n \frac{1}{k}}{\theta \sqrt{\sum_{k=1}^n \frac{1}{k} - \frac{\theta}{k^2}}} \approx \frac{\sum_{k=1}^n B_k(\theta) - \theta \log(n)}{\theta \sqrt{\log(n)}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathscr{N}(0,1)$$

Since we have a sum of Bernoulli random variables, one can expect a mod-Poisson convergence. Indeed, at the second order, setting $P_\gamma \sim \mathscr{P}(\gamma)$ with $\gamma := \theta \log(n)$, one gets

$$\begin{aligned} \frac{\mathbb{E}_\theta^{(n)}\left(x^C\right)}{\mathbb{E}\left(x^{P_\gamma}\right)} &= \frac{1}{e^{\theta \log n\,(x-1)}} \prod_{k=0}^{n-1} \frac{x\theta + k}{\theta + k} = \frac{1}{n^{\theta\,(x-1)}} \frac{\Gamma(x\theta + n - 1)}{\Gamma(x\theta)} \frac{\Gamma(\theta)}{\Gamma(\theta + n - 1)} \\ &\xrightarrow[n \to +\infty]{} \frac{\Gamma(\theta)}{\Gamma(x\theta)} \end{aligned}$$

using the well-known estimate $\Gamma(a+n)/\Gamma(b+n) \underset{n\to+\infty}{\sim} n^{a-b}$.

Setting $x = e^{i\alpha}$, one can prove that the convergence is locally uniform in $\alpha \in \mathbb{R}$ (see [68]). Remark that in the case of the uniform measure (i.e. $\theta = 1$), one has

$$\frac{\mathbb{E}_1^{(n)}\left(x^C\right)}{\mathbb{E}\left(x^{P_{\log(n)}}\right)} \xrightarrow[n\to+\infty]{} \Phi_C(x) := \frac{1}{\Gamma(x)}$$

Remark 3.3.1. The case of a sum of Bernoulli random variables was treated in full generality in (3.1), and the product form of the Fourier transform is equivalent to the following identity in distribution for the number of cycles C under the Ewens measure

$$\mathcal{L}_{\mathbb{P}_\theta^{(n)}}(C) = \sum_{k=1}^{n} B_k(\theta)$$

where the $(B_k(\theta))_k$ are independent Bernoulli random variables defined before. Setting $H_n(\theta) := \theta \sum_{k=1}^{n} \frac{1}{\theta+k-1}$, we hence have

$$\begin{aligned}\frac{\mathbb{E}_\theta^{(n)}\left(x^C\right)}{\mathbb{E}\left(x^{P_{H_n(\theta)}}\right)} &= \frac{1}{e^{\theta H_n(\theta)\,(x-1)}} \prod_{k=1}^{n}\left(1 + \frac{\theta}{\theta+k-1}(x-1)\right) \\ &\xrightarrow[n\to+\infty]{} \prod_{k\geqslant 1}\left(1 + \frac{\theta}{\theta+k-1}(x-1)\right) e^{-(x-1)\frac{\theta}{\theta+k-1}}\end{aligned}$$

In addition, $H_n(\theta) = \theta\left(\log n + \gamma + o(1)\right)$ whith γ the Euler-Mascheroni constant. Hence, we have the same result by replacing $H_n(\theta)$ by its approximation. This gives the product identity

$$\frac{\Gamma(\theta)}{\Gamma(x\theta)} = e^{-(x-1)\gamma\theta} \prod_{k\geqslant 1}\left(1 + \frac{\theta}{\theta+k-1}(x-1)\right) e^{-(x-1)\frac{\theta}{\theta+k-1}} \tag{3.3}$$

identity that has a certain probabilistic flavour since for $x = 0$ it expresses the equality in law

$$\mathrm{Gb}(1) \overset{\mathcal{L}}{=} -\gamma + \sum_{k\geqslant 1} \frac{\mathbb{e}_k - 1}{k} \tag{3.4}$$

between the Gumbel distribution of parameter 1 (i.e. $\mathbb{P}\left(\mathrm{Gb}(1) \leqslant x\right) = e^{-e^{-x}}$) and a sequence of i.i.d. exponential random variables of parameter 1 (i.e. $\mathbb{P}\left(\mathbb{e} \geqslant x\right) = e^{-x}$).

3.3.2 Random matrix theory

Let $U \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$ and consider the characteristic polynomial in 1, i.e. $Z_n := Z_U(0) = \det(U - I_n)$. All the computations done in this paragraph come from [19].

As seen in section 1.4.3, due to the decomposition (1.2.2), we have the equality in law (1.29)

$$\det\left(I_n - U\right) \overset{\mathcal{L}}{=} \prod_{k=1}^{n}\left(1 - e^{2i\pi U_k}\sqrt{\beta_{1,k-1}}\right)$$

Hence, the log of the characteristic polynomial taken in one point is a sum of independent random variables, and one has the Fourier transform (2.6). Write (with the usual determination of the log with a branch on $\mathbb{R}_-$)

$$\log(Z_n) \stackrel{\mathcal{L}}{=} \sum_{k=1}^{n} \log\left(1 - e^{2i\pi U_k}\sqrt{\beta_{1,k-1}}\right) =: \sum_{k=1}^{n} \ell_k$$

The classical first order renormalisation of this sum gives the CLT

$$\frac{\sum_{k=1}^n \ell_k - \mathbb{E}\left(\sum_{k=1}^n \ell_k\right)}{\sqrt{\operatorname{Var}\left(\sum_{k=1}^n \ell_k\right)}} \equiv \frac{\sum_{k=1}^n \ell_k}{\sqrt{\frac{1}{2}\log(n)}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathscr{N}_{\mathbb{C}}(0,1)$$

Let us focus on the real part of this logarithm, that is

$$L_n := \log|Z_n|$$

In terms of complex Fourier transform, one gets

$$\mathbb{E}\left(e^{zL_n/\sqrt{(\log n)/2}}\right) = \prod_{k=1}^{n} \frac{\Gamma(k)^2}{\Gamma\left(k+\frac{z}{2}\right)\Gamma\left(k-\frac{z}{2}\right)} = \mathbb{E}\left(e^{zG}\right)(1+o(1)) \quad \text{with} \quad G \sim \mathscr{N}(0,1)$$

The Keating-Snaith theorem can be stated as

$$\mathbb{E}\left(e^{2zL_n}\right) = n^{z^2}\widetilde{\Phi}_U(z)\,(1+o(1)) = e^{z^2\log(n)}\widetilde{\Phi}_U(z)\,(1+o(1)) \tag{3.5}$$

where $\widetilde{\Phi}_U$ is given in terms of Barnes' G-function by

$$\widetilde{\Phi}_U(z) := \frac{(\mathcal{G}(1+z))^2}{\mathcal{G}(1+2z)}$$

Since (c.f. (1.21)) for all $z \in \mathbb{C}$,

$$\mathcal{G}(z+1) := (2\pi)^{z/2} e^{-[(1+\gamma)z^2+z]/2} \prod_{n\geqslant 1}\left(1+\frac{z}{n}\right)^n e^{-z+(z^2/2n)}$$

with γ the Euler-Mascheroni constant and since the product is absolutely convergent, one can rewrite this limiting function as

$$\begin{aligned}
\widetilde{\Phi}_U(z) &= \frac{\left((2\pi)^{z/2}e^{-[(1+\gamma)z^2+z]/2}\right)^2}{(2\pi)^{2z/2}e^{-[(1+\gamma)(2z)^2+2z]/2}} \prod_{n\geqslant 1} \frac{\left(1+\frac{z}{n}\right)^{2n} e^{2(-z+(z^2/2n))}}{\left(1+\frac{2z}{n}\right)^n e^{-2z+((2z)^2/2n)}} \\
&= e^{(1+\gamma)z^2} \prod_{n\geqslant 1}\left(\frac{\left(1+\frac{z}{n}\right)^2}{1+\frac{2z}{n}}\right)^n e^{-z^2/n} \qquad (3.6)
\end{aligned}$$

Nevertheless, one can also use a different renormalising sequence $(\gamma_n)_n$ like $\gamma'_n = H_n$. In this case, one can write, using Fubini's theorem carefully for $z \in \mathbb{R}_+$ for example

$$\begin{aligned}
\mathbb{E}\left(e^{2zL_n}\right) &= \mathbb{E}\left(\prod_{k\leqslant n}\left|1-e^{2i\pi U_k}\sqrt{\beta_{1,k-1}}\right|^{2z}\right)\\
&= \prod_{k\leqslant n}\mathbb{E}\left(\left|1-e^{2i\pi U_k}\sqrt{\beta_{1,k-1}}\right|^{2z}\right)\\
&= \prod_{k\leqslant n}\mathbb{E}\left(\left|\sum_{m\geqslant 0}\frac{z^{\downarrow m}}{m!}(-1)^m e^{2i\pi U_k m}\beta_{1,k-1}^{m/2}\right|^2\right)\\
&= \prod_{k\leqslant n}\mathbb{E}\left(\sum_{m\geqslant 0}\left(\frac{z^{\downarrow m}}{m!}\right)^2\beta_{1,k-1}^m\right) \quad \text{and } \mathbb{E}\left(\beta_{1,k-1}^m\right) = \frac{m!}{k^{\uparrow m}}\\
&= \prod_{k\leqslant n}\left(\sum_{m\geqslant 0}\frac{\left(z^{\downarrow m}\right)^2}{m!\,k^{\uparrow m}}\right) = \prod_{k\leqslant n}\left(\sum_{m\geqslant 0}\frac{\left((-z)^{\uparrow m}\right)^2}{m!\,k^{\uparrow m}}\right)\\
&= \prod_{k\leqslant n}{}_2F_1\left(\begin{matrix}-z,-z\\ k\end{matrix}\,\middle|\,1\right)
\end{aligned}$$

Here, we have used the equivalent formulae

$$\begin{aligned}
(1-u)^{-z} &= \sum_{m\geqslant 0}\frac{z^{\uparrow m}}{m!}u^m := 1+\sum_{m\geqslant 1}\frac{z^{\uparrow m}}{m!}u^m\\
(1+u)^{z} &= \sum_{m\geqslant 0}\frac{z^{\downarrow m}}{m!}u^m := 1+\sum_{m\geqslant 1}\frac{z^{\downarrow m}}{m!}u^m
\end{aligned}$$

and the notations

$$\begin{aligned}
z^{\uparrow m} &:= z(z+1)\dots(z+m-1) = \frac{\Gamma(m+z)}{\Gamma(z)}\\
z^{\downarrow m} &:= z(z-1)\dots(z-m+1) = (-1)^m(-z)^{\uparrow m}\\
{}_2F_1\left(\begin{matrix}a,b\\ c\end{matrix}\,\middle|\,x\right) &:= \sum_{m\geqslant 0}\frac{a^{\uparrow m}\,b^{\uparrow m}}{c^{\uparrow m}}\frac{x^m}{m!} = \sum_{m\geqslant 0}\frac{a^{\uparrow m}\,b^{\uparrow m}}{c^{\uparrow m}\,1^{\uparrow m}}x^m \quad \text{(Gauss' hypergeometric function)}
\end{aligned}$$

Hence, with a locally uniform limit

$$\begin{aligned}
\frac{\mathbb{E}\left(e^{zL_n}\right)}{\mathbb{E}\left(e^{z\sqrt{H_n}G}\right)} &= \prod_{k\leqslant n}{}_2F_1\left(\begin{matrix}-z,-z\\ k\end{matrix}\,\middle|\,1\right)e^{-z^2/k}\\
&\xrightarrow[n\to+\infty]{} \prod_{k\geqslant 1}{}_2F_1\left(\begin{matrix}-z,-z\\ k\end{matrix}\,\middle|\,1\right)e^{-z^2/k} =: \Phi_U(z)
\end{aligned}\tag{3.7}$$

The equivalence between (3.6) and (3.7) comes from an identity of Gauss to express the value of a hypergeometric function in 1, that is

$${}_2F_1\left(\begin{matrix}a,b\\ c\end{matrix}\,\middle|\,1\right) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \quad \text{for } \mathfrak{Re}(c) > \mathfrak{Re}(a+b)$$

Note that we could deduce an equality in the same flavour as (3.3) by replacing the Gamma function by the Barnes' G-function and the exponential by the exponential of a square.

Remark 3.3.2. As one can see with the similarity of formulas between (3.7) and (3.1), the limiting function of a sum of independent random variables has always the same structure if the mod-* speed γ_n is the expectation of this sum : a converging product of mod-* renormalised Fourier transforms. Other examples will show that this is a general fact.

The two last examples were dealing with sums of independent random variables. We now present two examples with dependency.

3.3.3 Mod-Poisson convergence in probabilistic number theory

Let $\omega(N)$ denote the number of prime divisors of $N \in \mathbb{N}$, that is

$$\omega(N) := \sum_{p \in \mathcal{P}} \mathbb{1}_{\{p \triangleleft N\}}$$

One key theorem to understand the regularity of prime decomposition is the following :

Theorem 3.3.3 (Erdös-Kac)**.** *Let $U_n \sim \mathscr{U}\left([\![1, n]\!]\right)$. Then,*

$$\frac{\omega(U_n) - \log\log n}{\sqrt{\log\log n}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathscr{N}(0,1)$$

The intuition for this theorem is the following : since the rest of the division of U_n by 2 can only be 0 or 1, and since $\mathbb{P}\left(2 \triangleleft U_n\right) = [n/2]/n \to \frac{1}{2}$ when $n \to +\infty$, one could think of a *uniform measure on* $\mathbb{N}$ by inverting the probability and the limit into $\mathbb{P}\left(2 \triangleleft U_\infty\right) = \frac{1}{2}$. As remarked by Kac in [59], such a functional on set derived by taking the weak limit of $\mathscr{U}\left([\![1, n]\!]\right)$ is no more a σ-additive functional (i.e. a measure), only an additive functional. Hence, the probability axioms cannot apply.

Nevertheless, one can still do the approximation

$$\omega(U_n) = \sum_{p \in \mathcal{P}} \mathbb{1}_{\{p \triangleleft n\}} =: \sum_{p \in \mathcal{P}} B_p^{(n)} \underset{n \to +\infty}{\approx} \sum_{p \in \mathcal{P}, p \leqslant n} B_p^{(\infty)}$$

with the $B_p^{(\infty)}$'s are independent Bernoulli random variables defined by

$$\mathbb{P}\left(B_p^{(\infty)} = 1\right) = \frac{1}{p} = 1 - \mathbb{P}\left(B_p^{(\infty)} = 0\right)$$

To measure the accuracy of this last approximation (a sum of independent random variables), introduce the *independent model*

$$\Omega_n := \sum_{p \in \mathcal{P}, p \leqslant n} B_p^{(\infty)}$$

At the order of renormalisation of the CLT given by theorem 3.3.3, the independent model is accurate, since one can write

$$\frac{\omega(U_n) - \log\log n}{\sqrt{\log\log n}} \approx \frac{\Omega_n - \sum_{p \in \mathcal{P}, p \leqslant n} \frac{1}{p}}{\sqrt{\sum_{p \in \mathcal{P}, p \leqslant n} \frac{1}{p}}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathscr{N}(0,1)$$

Here, we have used the well-known estimate for the *prime harmonic sum*

$$H_n^{(\mathcal{P})} := \sum_{p\in\mathcal{P}, p\leqslant n} \frac{1}{p} = \log\log n + O(1) \tag{3.8}$$

Since Ω_n is a sum of $\{0,1\}$-Bernoulli random variables of square summable probabilities, the formula (3.1) gives

$$\mathbb{E}\left(z^{\Omega_n}\right) = \mathbb{E}\left(z^{P_{\gamma_n}}\right)\Phi_\Omega(z)\left(1+o(1)\right)$$

with $\gamma_n = H_n^{(\mathcal{P})}$ and

$$\Phi_\Omega(z) := \prod_{p\in\mathcal{P}}\left(1+\frac{z-1}{p}\right)e^{-\frac{z-1}{p}}$$

This model is interesting to understand the CLT, but it hides a certain amounts of information since at the second order of renormalisation the dependency re-appears : in the mod-Poisson setting, one has the following result due to Selberg (see [88])

$$\mathbb{E}\left(z^{\omega(U_n)}\right) = \mathbb{E}\left(z^{P_{\gamma_n}}\right)\widetilde{\Phi}_\omega(z)\left(1+O\left((\log n)^{\mathfrak{Re}(z-2)}\right)\right) \tag{3.9}$$

where, for $R>0$, the O is uniform for $|z|\leqslant R$, where $\gamma_n := \log\log(n)$ and

$$\widetilde{\Phi}_\omega(z) := \frac{1}{\Gamma(z)}\prod_{p\in\mathcal{P}}\left(1+\frac{z}{p-1}\right)\left(1-\frac{1}{p}\right)^z$$

Note that as a converging product of analytic functions, Φ_ω is analytic on $\mathbb{C}$. As the convergence holds on every compact of $\{z\in\mathbb{C}\,/\,\mathfrak{Re}(z)<1\}$, it holds locally uniformly in $z\in(0,1)$. It is moreover clear that for $x\geqslant 0$, $\Phi_\omega(x)\geqslant 0$ as $\Gamma(x)\geqslant 0$ and every term in the product is positive. This shows the mod-Poisson convergence.

One can write

$$\begin{aligned}\widetilde{\Phi}(z) &= \frac{1}{\Gamma(z)}\prod_{p\in\mathcal{P}}\left(1+\frac{z}{p-1}\right)\left(1-\frac{1}{p}\right)^z = \frac{1}{\Gamma(z)}\prod_{p\in\mathcal{P}}\frac{p}{p-1}\left(1+\frac{z-1}{p}\right)e^{z\log\left(1-\frac{1}{p}\right)}\\ &= \frac{1}{\Gamma(z)}\prod_{p\in\mathcal{P}}\left(1+\frac{z-1}{p}\right)e^{(z-1)\log\left(1-\frac{1}{p}\right)} = \frac{e^{-(z-1)\kappa_{\mathcal{P}}}}{\Gamma(z)}\prod_{p\in\mathcal{P}}\left(1+\frac{z-1}{p}\right)e^{-\frac{z-1}{p}}\end{aligned}$$

where $\kappa_{\mathcal{P}}$ is the positive constant given by

$$\kappa_{\mathcal{P}} := -\sum_{p\in\mathcal{P}}\left(\frac{1}{p}+\log\left(1-\frac{1}{p}\right)\right)\leqslant\frac{1}{2}\sum_{p\in\mathcal{P}}\frac{1}{p^2}<\infty \tag{3.10}$$

Hence, setting $\gamma_n' := \log\log n+\gamma-\kappa_{\mathcal{P}}$, one gets the mod-Poisson convergence at speed γ_n'

$$\mathbb{E}\left(z^{\omega(U_n)}\right) = \mathbb{E}\left(z^{P_{\gamma_n'}}\right)\Phi_\omega(z)\left(1+O\left((\log n)^{\mathfrak{Re}(z-2)}\right)\right)$$

with limiting function

$$\Phi_\omega(z) = \frac{e^{-(z-1)\gamma}}{\Gamma(z)} \prod_{p\in\mathcal{P}} \left(1 + \frac{z-1}{p}\right) e^{-\frac{z-1}{p}}$$

Last, using the formula (3.3), one has

$$\frac{1}{\Gamma(z)} = e^{(z-1)\gamma} \prod_{k\geqslant 1} \left(1 + \frac{z-1}{k}\right) e^{-\frac{z-1}{k}}$$

This gives the *splitting identity*

$$\Phi_\omega(z) = \Phi_C(z)\Phi_\Omega(z) \tag{3.11}$$

where Φ_Ω is the limiting function of the independent model Ω_n obtained at speed $\log\log n + \gamma + \kappa_{\mathcal{P}} = H_n^{(\mathcal{P})} + O(1)$ and Φ_C is the limiting function of is the number of cycles C of a random uniform permutation $\sigma \sim \mathrm{Haar}(\mathfrak{S}_n)$ obtained at speed $H_n = \sum_{k\leqslant n} \frac{1}{k}$. We can remark that

$$\begin{aligned}
\Phi_C(z) &= \prod_{k\in\mathbb{N}^*} \left(1 + \frac{z-1}{k}\right) e^{-\frac{z-1}{k}} \\
\Phi_\Omega(z) &= \prod_{p\in\mathcal{P}} \left(1 + \frac{z-1}{p}\right) e^{-\frac{z-1}{p}}
\end{aligned}$$

Hence, Φ_C and Φ_Ω have the same exact structure, except that they are characteristic of the set that indexes the sum of random variables in the independent model. This is due to the fact that the probabilities of the Bernoulli random variables involved are the same and only the sets indexing the probabilities differ.

As one can see on this example, the general phenomenon that is highlighted with the use of mod-Poisson convergence is a certain measure of lack of independence with respect to a natural independent model that explains a first order limit theorem (a CLT). This correction is not of independent additive nature since one cannot write $\omega(U_n)$ as the independent sum $\Omega_n + C(\sigma_n)$.

In the particular case of the mod-Poisson convergence with $\{0,1\}$-Bernoulli random variables of square summable probabilities as defined in example 3.2.6, a universality phenomenon seems to occur by the splitting into two models of the same $\{0,1\}$-Bernoulli type, as seen on the following example (see e.g. [68]) :

Example 3.3.4. Let $q = p^\nu$ and $\mathbb{F}_q$ denote the field with q elements. Denote by $\mathcal{P}(\mathbb{F}_q[X])$ the irreducible monic polynomials of $\mathbb{F}_q[X]$ and by $\omega_q(P_n)$ the number of divisors of P_n defined, for all monic $Q \in \mathcal{P}(\mathbb{F}_q[X])$ by

$$\omega_q(Q) := \sum_{\pi\in\mathcal{P}(\mathbb{F}_q[X])} \mathbb{1}_{\{\pi \triangleleft Q\}}$$

Let Q_n be a random monic polynomial of degree less than n selected according to the uniform measure of this finite set. It is shown in [68] that[1]

$$\frac{\mathbb{E}\left(z^{\omega_q(Q_n)}\right)}{\mathbb{E}\left(z^{P_{\gamma_n}}\right)} = \Phi_{\omega_q}(z)(1+o(1))$$

where $\gamma_n = H_n + \kappa_q$, where κ_q is defined by

$$\kappa_q := -\sum_{\pi\in\mathcal{P}(\mathbb{F}_q[X])}\left(\frac{1}{|\pi|_q} + \log\left(1-\frac{1}{|\pi|_q}\right)\right) < \infty$$

with $|\pi|_q := q^{\deg(\pi)}$, and where

$$\Phi_{\omega_q}(z) = \frac{e^{-(z-1)\gamma}}{\Gamma(z)}\prod_{\pi\in\mathcal{P}(\mathbb{F}_q[X])}\left(1+\frac{z-1}{|\pi|_q}\right)e^{-\frac{z-1}{|\pi|_q}} =: \Phi_C(z)\Phi_{\Omega_q}(z)$$

This form is reminiscent of (3.11), with a corrective model given by $C(\sigma_n)$ for $\sigma_n \sim \mathrm{Haar}(\mathfrak{S}_n)$ and an independent model given by

$$\Omega_{q,n} := \sum_{\substack{\deg(\pi)\leqslant n\\ \pi\in\mathcal{P}(\mathbb{F}_q[X])}} B_\pi = \sum_{d=1}^{n}\sum_{\substack{\deg(\pi)=d\\ \pi\in\mathcal{P}(\mathbb{F}_q[X])}} B_\pi$$

where the $(B_\pi)_\pi$ are independent Bernoulli random variables such that $\mathbb{P}\left(B_\pi = 1\right) = \frac{1}{|\pi|_q} = 1 - \mathbb{P}\left(B_\pi = 0\right)$.

Note that using Moebius inversion, one has

$$c_d^{(q)} := \sum_{\pi\in\mathcal{P}(\mathbb{F}_q[X])}\mathbb{1}_{\{\deg(\pi)=d\}} = \frac{1}{d}\sum_{\ell\triangleleft d}\mu(\ell)q^{d/\ell} = \frac{q^d}{d} + O\left(q^{d/2}\right)$$

hence one can write

$$\Phi_{\Omega_q}(z) := \prod_{\pi\in\mathcal{P}(\mathbb{F}_q[X])}\left(1+\frac{z-1}{|\pi|_q}\right)e^{-\frac{z-1}{|\pi|_q}} = \prod_{d\geqslant 1}\left(\left(1+\frac{z-1}{q^d}\right)e^{-\frac{z-1}{q^d}}\right)^{c_d^{(q)}}$$

Φ_{Ω_q} is the same type of limiting function as Φ_Ω, but this time the indexing set is the multiset $\{\{\,|\pi|_q\,, \pi \in \mathcal{P}(\mathbb{F}_q[X])\,\}\}$ (this amounts to have an independent model with a sum independent random walks).

What one can see on this last example is an avatar of a universality phenomenon for $\{0,1\}$-Bernoulli structures. This would be interesting to study more general class of models coming from general divisibility structures (for example $\mathbb{Z}[i]$, $\mathbb{F}_q[X,Y]$ or graphs) and more generally the most natural completely dependent model constructed from uniformly bounded orthogonal bases in the flavour of theorem 3.1.2, with indicator functions depending on the same uniform distribution.

[1] The formula used is in the form of (3.9), but the same manipulations as before allow to write it with the given form.

Remark 3.3.5. A possible interpretation of Φ_ω in the framework of multisets is possible since

$$\Phi_\omega(z) = \prod_{k\in\mathbb{N}^*}\left(1+\frac{z-1}{k}\right)e^{-\frac{z-1}{k}}\prod_{p\in\mathcal{P}}\left(1+\frac{z-1}{p}\right)e^{-\frac{z-1}{p}} = \prod_{k\in\mathbb{N}^*}\left[\left(1+\frac{z-1}{k}\right)e^{-\frac{z-1}{k}}\right]^{1+\mathbb{1}_{\{k\in\mathcal{P}\}}}$$

i.e. Φ_ω is the function associated to the multiset $\{\mathbb{N}^*, \mathcal{P}\}$ where the primes are repeated.

3.3.4 Mod-Gaussian convergence in probabilistic number theory : the moments conjecture

A "multiplication pardigm" has emerged from the last examples. It asserts that the limiting mod-* function of a dependant sequence of random variables satisfying a CLT is the product of two limiting mod-* functions associated 1) to a natural independent model and 2) to a corrective model also built from a sum of independent random variables which is characteristic of a more general class of models (a class of *universality*).

In the case of the Riemann Zeta function on the critical axis, one can write formally (since the product is not convergent on this axis)

$$\left|\zeta\left(\frac{1}{2}+iTU\right)\right| \equiv \prod_{p\in\mathcal{P}}\left|1-\frac{1}{p^{1/2+iTU}}\right|^{-1} = \prod_{p\in\mathcal{P}}\left|1-\frac{e^{-iTU\log(p)}}{\sqrt{p}}\right|^{-1}$$

In fact, considering the following truncation for any $\varepsilon > 0$

$$\zeta_T^{(\varepsilon)} := \prod_{\substack{p\in\mathcal{P}\\ p\leqslant T^\varepsilon}}\left|1-\frac{e^{-iTU\log(p)}}{\sqrt{p}}\right|^{-1}$$

we are exactly in the case of the "totally dependant" CLT of Salem-Zygmund enunciated in theorem 3.1.2, more precisely in the case of linearly independent periods considered in [59], since the $\log(p)$'s are linearly independent. Note that $\log\zeta_T^{(\varepsilon)}$ is a sum of logarithms of trigonometric functions, and that a second truncation of their Fourier development would exactly give the Salem-Zygmund case ; such a sum was exactly considered by Selberg in his original proof (see for instance [97], part 4).

Historically, first attempts at computating the Fourier transform of $\log\left|\zeta\left(\frac{1}{2}+iTU\right)\right|$ arose from the computation of integers moments of $\left|\zeta\left(\frac{1}{2}+iTU\right)\right|$. The first result of this type is due to Hardy and Littlewood who showed that

$$\mathbb{E}\left(\left|\zeta\left(\frac{1}{2}+iTU\right)\right|\right) = \int_0^1\left|\zeta\left(\frac{1}{2}+iTv\right)\right|dv \underset{T\to+\infty}{\sim} \log T$$

followed soon by Ingham who obtained

$$\mathbb{E}\left(\left|\zeta\left(\frac{1}{2}+iTU\right)\right|^2\right) = \int_0^1\left|\zeta\left(\frac{1}{2}+iTv\right)\right|^2 dv \underset{T\to+\infty}{\sim} \frac{(\log T)^4}{2\pi^2}$$

Nothing else is known for higher order moments, but a first conjecture due to Conrey and Gosh (see [28]) and a second conjecture due to Conrey and Gonek (see [27]) assert that

$$\mathbb{E}\left(\left|\zeta\left(\frac{1}{2}+iTU\right)\right|^{2k}\right) = \int_0^1 \left|\zeta\left(\frac{1}{2}+iTv\right)\right|^{2k} dv \underset{T\to+\infty}{\sim} a_k u_k (\log T)^{k^2}$$

with

$$a_k = \prod_{p\in\mathcal{P}} {}_2F_1\left(\begin{matrix}k,k\\1\end{matrix}\middle|\frac{1}{p}\right) e^{k^2\log\left(1-\frac{1}{p}\right)} = e^{\kappa_{\mathcal{P}}k^2}\prod_{p\in\mathcal{P}} {}_2F_1\left(\begin{matrix}k,k\\1\end{matrix}\middle|\frac{1}{p}\right) e^{k^2/p}$$

where $\kappa_{\mathcal{P}}$ is defined in (3.10), and

$$u_3 = \frac{42}{9!}, \qquad u_4 = \frac{24024}{16!}$$

The hardest part in this conjecture was the computation of these latest coefficients. They appear as an approximation of the ζ product for $p \geqslant X = O\left((\log T)^{2-\varepsilon}\right)$, the first term corresponding to the truncation of the product for $p \leqslant X$. Using this philosophy, Keating and Snaith were able to produce the following conjecture

Conjecture 3.3.6 (Keating-Snaith, [62])**.** The following equality holds for all $k \geqslant 1$

$$\mathbb{E}\left(\left|\zeta\left(\frac{1}{2}+iTU\right)\right|^{2k}\right) \underset{T\to+\infty}{\sim} \Phi_U(k)\Phi_A(k)(e^{\kappa_{\mathcal{P}}}\log T)^{k^2}$$

where

$$\begin{aligned}\Phi_A(z) &:= \prod_{p\in\mathcal{P}} {}_2F_1\left(\begin{matrix}z,z\\1\end{matrix}\middle|\frac{1}{p}\right) e^{-z^2/p} \\ \Phi_U(z) &:= \prod_{k\in\mathbb{N}^*} {}_2F_1\left(\begin{matrix}-z,-z\\k\end{matrix}\middle|1\right) e^{-z^2/k}\end{aligned} \tag{3.12}$$

In particular,

$$\Phi_U(k) = \prod_{\ell=1}^{k} \frac{\ell!}{(\ell+k)!}$$

By analogy with the mod-Poisson case, the origin of this conjecture is clear : the splitting into an independent model and a corrective model, the first one characteristic of the CLT encountered, the second one being a *universal* [2] term involving only the fluctuations of the zeros of the Zeta function, hence that should be the same for each function whose zeros have the same fluctuations, which is the case of a random matrix $U \sim \mathrm{Haar}(\mathcal{U}_n(\mathbb{C}))$ in virtue of the Montgommery and Dyson theorems. Note that another model involving the sin kernel

[2] The universality is a concept that comes initially from the study of critical models in mechanical statistics and that extended to the whole probability theory. It amounts to say that a large class of objects define the same limiting model after a suitable renormalisation. It is for example the case with the Gaussian law and a large class of sequences of random variables, or the Brownian motion and another class of random variables, in accordance to Donsker's invariance principle.

is perfectly adequate and this factor is more likely to be called the "universal factor" rather than "matrix factor" or "combinatorial factor". Note also that the splitting into a first part consisting into primes and a second part consisting into zeros is at the core of Selberg's proof of his CLT (see [97] for an informal presentation).

Remark 3.3.7. The Keating-Snaith philosophy is twofold : use the characteristic polynomial on the circle to produce conjectures in number theory since the computations are notoriously harder to achieve in number fields, or take number-theoretic results on L-functions to check that the same results hold for characteristic polynomials on the circle. Moreover, generalisations to random permutations or function fields can enter into this scheme of comprehension before reaching number fields.

This philosophy can be roughly sumarized into the following diagram

$$\begin{array}{ccc}
\text{Random permutations} & & Z_\sigma(\theta) = \det\left(I - e^{i\theta}\sigma\right) \\
\downarrow & & \downarrow \\
\text{Random matrices} & & Z_U(\theta) = \det\left(I - e^{i\theta}U\right) \\
\downarrow & & \downarrow \\
\text{Function fields} & & \zeta_{\mathbb{F}_p[X]}\left(\frac{1}{2} + iTU\right) \\
\downarrow & & \downarrow \\
\text{Number fields} & & \zeta\left(\frac{1}{2} + iTU\right)
\end{array}$$

A refinement of the latest heuristic is made by the Gonek-Hughes-Keating *hybrid product* (see [46]). It consists in factorising the product of $\zeta\left(\frac{1}{2} + iTU\right)$ into the two parts [3]

$$\zeta\left(\frac{1}{2} + iTU\right) = P_X\left(\frac{1}{2} + iTU\right) Z_X\left(\frac{1}{2} + iTU\right) := \prod_{\substack{p \leqslant X \\ p \in \mathcal{P}}} \left(1 - \frac{p^{-it}}{\sqrt{p}}\right)^{-1} \prod_{\substack{p > X \\ p \in \mathcal{P}}} \left(1 - \frac{p^{-it}}{\sqrt{p}}\right)^{-1}$$

and make the following *splitting conjecture*

$$\mathbb{E}\left(\left|\zeta\left(\frac{1}{2} + iTU\right)\right|^{2k}\right) \underset{T\to+\infty}{\sim} \mathbb{E}\left(\left|P_X\left(\frac{1}{2} + iTU\right)\right|^{2k}\right) \mathbb{E}\left(\left|Z_X\left(\frac{1}{2} + iTU\right)\right|^{2k}\right)$$

They prove that for $X = O\left((\log T)^{2-\varepsilon}\right)$

$$\mathbb{E}\left(\left|P_X\left(\frac{1}{2} + iTU\right)\right|^{2k}\right) \underset{T\to+\infty}{\sim} \Phi_A(k) \left(e^\gamma \log X\right)^{k^2}$$

[3] This is an approximation of the real decomposition. For the precise definition of Z_X and P_X, see [46].

They conjecture that for the same value of X

$$\mathbb{E}\left(\left|Z_X\left(\frac{1}{2}+iTU\right)\right|^{2k}\right)\underset{T\to+\infty}{\sim}\Phi_U(k)\left(e^{-\gamma}\frac{\log T}{\log X}\right)^{k^2}$$

Here again, as in the proof of Selberg's CLT, the dichotomy between small primes and big primes governed by the fluctuations of the zeros is apparent. An analogous explanation in the mod-Poisson case for $\omega(U_n)$ can be made since $\omega(U_n)=\sum_{p\in\mathcal{P}}f_p(U_n)$ with $f_p(x):=\mathbb{1}_{\{p\mid x\}}$ and $f_p(x)=-\log\left|1-e^{2i\pi x\log(p)}/\sqrt{p}\right|$ in the case of $\log|\zeta(1/2+2i\pi TU)|$: the source of the splitting comes from the separation into big primes corresponding to universal components behaving like big cycles of $\sigma\sim\mathrm{Haar}(\mathfrak{S}_n)$, and small primes that give the specificity of the mod-* limiting function, but that are universal at the order of renormalisation of the CLT (i.e. that belong to the class of universality of the Gaussian distribution).

We remark moreover that the limiting functions given in (3.12) and (3.7) are of the same type. Indeed, one has for the independent model and its correction, with i.i.d. uniform and β random variables (each independent)

$$\begin{aligned}\zeta_n^* &:= \prod_{\substack{p\leqslant n\\ p\in\mathcal{P}}}\left|1-e^{2i\pi U_p}\frac{1}{\sqrt{p}}\right|^{-1}\\ |Z_n| &:= \prod_{k\leqslant n}\left|1-e^{2i\pi U_k}\sqrt{\beta_{1,k-1}}\right|\end{aligned}$$

One important feature of the two formulas is the presence of the uniform random variables on the unit circle. One can also notice the similarity of the forms of the random variables occuring. In the same vein as the computation of Φ_U, one can write (see [19])

$$\begin{aligned}\mathbb{E}\left(e^{2z\log\zeta_n^*}\right) &= \mathbb{E}\left(\prod_{\substack{p\leqslant n\\ p\in\mathcal{P}}}\left|1-e^{2i\pi U_p}\sqrt{p^{-1}}\right|^{-2z}\right)\\ &= \prod_{\substack{p\leqslant n\\ p\in\mathcal{P}}}\mathbb{E}\left(\left|1-e^{2i\pi U_k}\sqrt{p^{-1}}\right|^{-2z}\right)\\ &= \prod_{\substack{p\leqslant n\\ p\in\mathcal{P}}}\mathbb{E}\left(\left|\sum_{m\geqslant 0}\frac{z^{\uparrow m}}{m!}e^{2i\pi U_k m}p^{-m/2}\right|^2\right)\\ &= \prod_{\substack{p\leqslant n\\ p\in\mathcal{P}}}\left(\sum_{m\geqslant 0}\left(\frac{z^{\uparrow m}}{m!}\right)^2 p^{-m}\right)=\prod_{\substack{p\leqslant n\\ p\in\mathcal{P}}}\left(\sum_{m\geqslant 0}\frac{\left(z^{\uparrow m}\right)^2}{m!}\frac{p^{-m}}{m!}\right)\\ &= \prod_{\substack{p\leqslant n\\ p\in\mathcal{P}}}{}_2F_1\left(\begin{matrix}z,z\\1\end{matrix}\middle|p^{-1}\right)\end{aligned}$$

Hence, with a locally uniform limit

$$\begin{aligned}\frac{\mathbb{E}\left(e^{z\log\zeta_n^*}\right)}{\mathbb{E}\left(e^{z\sqrt{H_n^{(\mathcal{P})}}G}\right)} &= \prod_{\substack{p\leqslant n\\ p\in\mathcal{P}}} {}_2F_1\left(\begin{matrix} z,z\\ 1\end{matrix}\middle| p^{-1}\right)e^{-z^2/p}\\ &\xrightarrow[n\to+\infty]{} \prod_{p\in\mathcal{P}} {}_2F_1\left(\begin{matrix} z,z\\ 1\end{matrix}\middle| p^{-1}\right)e^{-z^2/p} =: \Phi_A(z)\end{aligned}$$

Last, for $k\geqslant 1$

$$\mathbb{E}\left(\beta_{1,k-1}\right)=\frac{1}{k},\qquad \operatorname{Var}\left(\beta_{1,k-1}\right)=\frac{1}{k^2}\frac{k-1}{k+1}=:\frac{1}{k^2}(1+\varepsilon_k)$$

and for all $x>0$

$$\mathbb{P}\left(|\beta_{1,k-1}-1/k|\geqslant x\right)\leqslant\frac{1}{k^2x^2}(1+\varepsilon_k)$$

i.e. for k large, the random variable $\beta_{1,k-1}$ is close to $1/k$ with a sufficiently high probability, so one can think of the corrective model as a slight random perturbation of the model given by

$$Z_n^*=\prod_{k\leqslant n}\left|1-e^{2i\pi U_k}\sqrt{k^{-1}}\right|$$

which completes the analogy with the mod-Poisson case.

3.3.5 A summary

The last examples revealed that mod-* convergence is characteristic of a second level of universality. As a first level is concerned with the convergence in distribution, for instance Gaussian or Poisson, a second level highlights the rôle of a particular additional model that corrects a first intuitive model characteristic of the convergence in distribution, at least in the case of the precedent examples where an independent model in "naturally" present[4]. This corrective model is universal and shared by a large class of converging sequences. It can be tought of as a subclass of universality of the distribution for which the sequence converges in the mod sense, i.e. a *second-order universality.*

Our goal is now to give a possible probabilistic interpretation of mod-* convergence and explain the nature of this limiting function Φ.

[4]Note that there are examples where a conjectural independent model is yet to be found ; random variables constructed from arithmetic considerations thus appear as extremely special. An explanation of the existence of a natural independent model for $\omega(U_n)$ will be given in section 3.6.

3.4 A possible probabilistic interpretation of mod-* convergence

3.4.1 Classical biases and changes of probability

A fundamental operation in probability theory is the change of probability by means of a weight on the initial probability measure. This weight is called *bias* or *penalisation* and we will use undifferently both terminology (for references, see any classical book on probability theory, for instance [38]).

Definition 3.4.1 (Bias/penalisation of measure)**.** Let X be a real random variable in the probability space endowed with the measure $\mathbb{P}$ and denote by $\mathbb{P}_X$ its law, i.e. if A is a measurable set, $\mathbb{P}_X(A) := \mathbb{P}(X \in A)$. For $f \in L^1(\mathbb{P}_X)$, $f \geqslant 0$, the penalisation (or bias) of $\mathbb{P}_X$ by f is the probability measure $\mathbb{P}_Y$ denoted by

$$\mathbb{P}_Y := \frac{f(X)}{\mathbb{E}\left(f(X)\right)} \bullet \mathbb{P}_X$$

This definition is equivalent to the following : for all $g \in L^\infty(\mathbb{P}_X)$,

$$\mathbb{E}\left(g(Y)\right) = \frac{\mathbb{E}\left(f(X)g(X)\right)}{\mathbb{E}\left(f(X)\right)}$$

Classical bias in probability theory allow to understand "pathwise transformations" induced by such a transformation.

Example 3.4.2. The most classical change of probability concerns the passage from $\mathscr{N}(0,1)$ to $\mathscr{N}(\mu,1) \stackrel{\mathcal{L}}{=} \mu + \mathscr{N}(0,1)$. Indeed, if $X \sim \mathscr{N}(0,1)$, one easily checks that

$$\mathbb{P}_{X+\mu} = \frac{e^{\mu X}}{\mathbb{E}\left(e^{\mu X}\right)} \bullet \mathbb{P}_X = e^{\mu X - \mu^2/2} \bullet \mathbb{P}_X \tag{3.13}$$

A tensorisation of this identity for an infinite number of i.i.d. Gaussian random variables gives the celebrated Girsanov theorem in the brownian setting

$$\mathbb{P}_{X+\mu\langle X,X\rangle}\big|_{\mathcal{F}_t} = e^{\mu X_t - \mu^2\langle X,X\rangle_t/2} \bullet \mathbb{P}_X\big|_{\mathcal{F}_t}$$

where $\mathbb{P}_X$ is the Wiener measure, X_t the canonical process defined by $X_t(\omega) = \omega(t)$, $(\mathcal{F}_t)_t$ the canonical filtration defined by $\mathcal{F}_t = \sigma(X_s, s \leqslant t)$ and $\langle X, X\rangle_t = t$ here.

Hence, an exponential bias is equivalent to a translation of the canonical evaluation (resp. the canonical process) in the Gaussian (resp. Brownian) setting.

A classical transform in probability theory is made with the weight $x \mapsto x$ when the random variable is positive.

Definition 3.4.3. Let $X \geqslant 0$ be a random variable with expectation $\mu := \mathbb{E}(X) < \infty$. A random variable $X^{(s)}$ is said to be a *size-bias transform* of X if, for all real functions f such that $\mathbb{E}\left(|Xf(X)|\right) < \infty$

$$\mathbb{E}\left(Xf(X)\right) = \mu\mathbb{E}\left(f\left(X^{(s)}\right)\right)$$

An equivalent definition is thus

$$\mathbb{P}_{X^{(s)}} := \frac{X}{\mathbb{E}\left(X\right)} \bullet \mathbb{P}_X \tag{3.14}$$

Example 3.4.4. A classical change of measure for a random walk with positive increments is given by its size-bias coupling, i.e. given $(X_k)_k$ a sequence of i.i.d. positive random variables of expectation $\mathbb{E}(X_k) := 1$ defined on the same probability space, the random walk $(S_n)_n$ of increments $(X_k)_k$ is given by

$$S_n := \sum_{k=1}^{n} X_k$$

The size-bias transform of S_n is the random variable $S_n^{(s)}$ of law given by

$$\mathbb{P}_{S_n^{(s)}} := \frac{S_n}{n} \bullet \mathbb{P}_{S_n}$$

A pathwise construction of such a random variable is given by the following

Lemma 3.4.5 (Size-bias coupling of an independent sum, [2]). *Let $(Y_k)_k$ be a sequence of independent positive integrable random variables, independent of $(X_k)_k$ and having the same distribution as $(X_k)_k$ and let $I \in [\![1, n]\!]$ be a random index independent of $(X_k)_k$ and $(Y_k)_k$ of law given by*

$$\mathbb{P}(I = k) = \frac{\mathbb{E}(X_k)}{\sum_{\ell=1}^{n} \mathbb{E}(X_\ell)}$$

Then,

$$S_n^{(s)} \stackrel{\mathcal{L}}{=} S_n - X_I + Y_I^{(s)}$$

and in particular, if $(Y_k)_k$ is defined on the same probability space as $(X_k)_k$, one has a natural coupling $\left(S_n, S_n^{(s)}\right)$.

Proof. Let f be a bounded function and $S_n^{\langle -k \rangle} := \sum_{\ell \neq k} X_\ell$. Then, by independence,

$$\begin{aligned}
\mathbb{E}\left(f\left(S_n^{(s)}\right)\right) &:= \frac{1}{\mathbb{E}(S_n)} \mathbb{E}(S_n f(S_n)) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}(X_k f(S_n)) \\
&= \frac{1}{\mathbb{E}(S_n)} \sum_{k=1}^{n} \mathbb{E}\left(X_k f\left(S_n^{\langle -k \rangle} + X_k\right)\right) \\
&= \frac{1}{\mathbb{E}(S_n)} \sum_{k=1}^{n} \mathbb{E}(X_k)\, \mathbb{E}\left(f\left(S_n^{\langle -k \rangle} + Y_k^{(s)}\right)\right) \\
&= \mathbb{E}\left(f\left(S_n^{\langle -I \rangle} + Y_I^{(s)}\right)\right) = \mathbb{E}\left(f\left(S_n - X_I + Y_I^{(s)}\right)\right)
\end{aligned}$$

□

A last type of useful bias concerns the discrete equivalent of Girsanov's theorem, using the discrete Bernouilli random walk instead of the Brownian motion.

Lemma 3.4.6 (Exponential bias of a random walk)**.** *Let $(B_k)_k$ be a sequence of independent $\{0,1\}$-Bernoulli random variables, each of probability p_k to be equal to 1. Define, for a certain $x > 0$,*

$$
\begin{aligned}
S_n &:= \sum_{k=1}^{n} B_k \\
\mathbb{P}_{S_n(x)} &:= \frac{x^{S_n}}{\mathbb{E}\left(x^{S_n}\right)} \bullet \mathbb{P}_{S_n}
\end{aligned}
$$

Then,

$$
S_n(x) \stackrel{\mathcal{L}}{=} \sum_{k=1}^{n} B_k(x)
$$

with $\mathbb{P}\left(B_k(x) = 1\right) = \frac{xp_k}{xp_k + 1 - p_k} = 1 - \mathbb{P}\left(B_k(x) = 0\right)$.

Proof. Let $y > 0$. Then, by independence,

$$
\mathbb{E}\left(y^{S_n(x)}\right) := \frac{\mathbb{E}\left((xy)^{S_n}\right)}{\mathbb{E}\left(x^{S_n}\right)} = \prod_{k=1}^{n} \frac{\mathbb{E}\left((xy)^{B_k}\right)}{\mathbb{E}\left(x^{B_k}\right)} = \prod_{k=1}^{n} \left(\frac{p_k xy + 1 - p_k}{p_k x + 1 - p_k}\right) = \prod_{k=1}^{n} \mathbb{E}\left(y^{B_k(x)}\right)
$$

□

3.4.2 Bias and mod-* convergence

One important property of the operation of penalising is its associativity : if f and g are two bounded positive functions (say), one has

$$
\frac{(fg)(X)}{\mathbb{E}\left((fg)(X)\right)} \bullet \mathbb{P}_X = \frac{f(Y)}{\mathbb{E}\left(f(Y)\right)} \bullet \mathbb{P}_Y \quad \text{with} \quad \mathbb{P}_Y = \frac{g(X)}{\mathbb{E}\left(g(X)\right)} \bullet \mathbb{P}_X
$$

i.e. with the classical convention of the canonical evaluation $(X(\omega) = \omega)$

$$
\frac{(fg)(X)}{\mathbb{E}\left((fg)(X)\right)} \bullet \mathbb{P}_X = \frac{f(X)}{\mathbb{E}\left(f(X)\right)} \bullet \left(\frac{g(X)}{\mathbb{E}\left(g(X)\right)} \bullet \mathbb{P}_X\right)
$$

We remark that this transitivity/associativity property depends on the normalisation, hence is not linear in f. This is also the case for the product of two mod-* limiting functions : it is is again a mod-* limiting function, but with a change of renormalisation, i.e. if

$$
\begin{aligned}
(X_n, \gamma_n)_n &\xrightarrow[n \to +\infty]{\text{mod-P}} \Phi_1 \\
(Y_n, \gamma'_n)_n &\xrightarrow[n \to +\infty]{\text{mod-P}} \Phi_2
\end{aligned}
$$

then, $\Phi_1 \Phi_2$ is again a limiting mod-* function for $X_n + Y_n$ (supposing that they are on the same probability space), and with parameter $\gamma_n + \gamma'_n$.

It seems thus reasonable to interpret such a convergence in termes of bias. And indeed, one has the

Theorem 3.4.7 (A possible probabilistic interpretation of mod-gaussian convergence)**.** *Let $(\gamma_n)_n$ be a real sequence such that $\gamma_n \xrightarrow[n\to+\infty]{} +\infty$ and Φ be an admissible function for the mod-gaussian convergence (i.e. a continuous complex function satisfying $\Phi(0) = 1$ and $\overline{\Phi(u)} = \Phi(-u)$).*

Suppose moreover that

1. *Φ can be analytically extended on the whole complex plane and satisfies, $\forall\, \beta \in \mathbb{R}$,*

$$\begin{aligned} \sup_{z\in a+i[0,\beta]} |\Phi(z)| &< \infty \quad \forall\, a \in \mathbb{R} \\ \sup_{z\in a+i[0,\beta]} |\Phi(z)| &\xrightarrow[a\to\pm\infty]{} 0 \end{aligned} \tag{3.15}$$

2. *$\Phi(ix) = \Phi(x)$ for all $x \in \mathbb{R}$,*
3. *$\Phi(x) \geqslant 0$ for all $x \in \mathbb{R}$,*

Define the distribution $\mathbb{P}_{\mathcal{H}_n}$ of a random variable $\mathcal{H}_n$ by the following penalisation

$$\mathbb{P}_{\mathcal{H}_n} := \frac{\Phi\left(\frac{G}{\gamma_n}\right)}{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n}\right)\right)} \bullet \mathbb{P}_{\gamma_n G} \tag{3.16}$$

Then,

$$(\mathcal{H}_n, \gamma_n) \xrightarrow[n\to\infty]{\text{mod-G}} \Phi$$

Example 3.4.8. Before proving the theorem, we can give an example of such a function Φ, that will be our guiding example. For $C > 0$, set

$$\Phi(x) = e^{-Cx^4}$$

This function is the mod-gaussian limit of (a renormalisation of)

$$S_n := \sum_{k=1}^{n} X_k$$

with $(X_k)_k$ a sequence of i.i.d. symmetric random variables (that is $X \overset{\mathcal{L}}{=} -X$) having a moment of order 4. To prove that $(n^{-1/4}S_n, n^{1/4})_n$ converges in the mod-gaussian sense to such a function Φ with parameter $C = (3 - \mathbb{E}\left(X^4\right))/24$, we can suppose that $\mathbb{E}\left(X^2\right) = 1$ and we suppose that $\kappa := \mathbb{E}\left(X^4\right) < 3$. Then, setting $\phi_X(x) := \mathbb{E}\left(e^{ixX}\right)$, we have

$$\begin{aligned} \mathbb{E}\left(e^{ixn^{-1/4}S_n}\right) &= \mathbb{E}\left(e^{ix\sum_{k=1}^{n} X_k/n^{1/4}}\right) = \left(\mathbb{E}\left(e^{ixX/n^{1/4}}\right)\right)^n \\ &= e^{n\log\left(\phi_X(x/n^{1/4})\right)} \\ &= e^{n\log\left(\phi_X(0)+\phi''_X(0)\frac{x^2}{2\sqrt{n}}+\phi_X^{(4)}(0)\frac{x^4}{24n}+o_x\left(\frac{1}{n}\right)\right)} \\ &= e^{n\log\left(1+\frac{x^2}{2\sqrt{n}}+\kappa\frac{x^4}{24n}+\frac{x^4}{n}\varepsilon_1\left(\frac{x}{n^{1/4}}\right)\right)} \\ &= e^{n\left(\frac{x^2}{2\sqrt{n}}+\kappa\frac{x^4}{24n}-\frac{1}{2}\left(\frac{x^2}{2\sqrt{n}}\right)^2+\frac{x^4}{n}\varepsilon_2\left(\frac{x}{n^{1/4}}\right)\right)} \\ &= e^{\sqrt{n}\frac{x^2}{2}+(\kappa-3)\frac{x^4}{24}+\frac{x^4}{n}\varepsilon_2\left(\frac{x}{n^{1/4}}\right)} \end{aligned}$$

Here, ε_1 and ε_2 are functions that tend to 0 in 0 and are bounded on a compact neighborhood of 0. Hence, we can consider the following convergence that holds locally uniform in x in a certain interval around 0

$$\frac{\mathbb{E}\left(e^{ixn^{-1/4}S_n}\right)}{\mathbb{E}\left(e^{ixn^{1/4}G}\right)} \xrightarrow[n\to+\infty]{} e^{-\frac{(3-\kappa)}{24}x^4}$$

We can check moreover that the required hypotheses of analyticity and boundedness in a horizontal band are fullfilled, that is

$$\sup_{z\in a+i[0,\beta]}\left|e^{-Cz^4}\right| = \sup_{y\in[0,\beta]}\left|e^{-C(a+iy)^4}\right| = \sup_{y\in[0,\beta]} e^{-C(a^4+y^4-6a^2y^2)} \leqslant C'_\beta e^{-Ca^4/2} \xrightarrow[a\to\pm\infty]{} 0$$

Thus, according to the theorem, the random variable $\mathcal{H}_n$ of distribution given by (3.16) with $\Phi_C(x) = e^{-Cx^4}$ "looks like" the distribution of $n^{-1/4}S_n$, in the same vein as the distribution of S_n "looks like" the distribution of $G \sim \mathscr{N}(0,1)$ according to the Central Limit Theorem.

Now, we prove theorem 3.4.7.

Proof. For $\theta\in\mathbb{R}$, write

$$\frac{\mathbb{E}\left(e^{i\theta\mathcal{H}_n}\right)}{\mathbb{E}\left(e^{i\theta\gamma_nG}\right)} = \frac{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n}\right)e^{i\theta\gamma_nG}\right)}{\mathbb{E}\left(e^{i\theta\gamma_nG}\right)\mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n}\right)\right)} = \frac{\mathbb{E}\left(\frac{e^{i\theta\gamma_nG}}{\mathbb{E}\left(e^{i\theta\gamma_nG}\right)}\Phi\left(\frac{G}{\gamma_n}\right)\right)}{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n}\right)\right)} =: \frac{\int_{\mathbb{R}}\Phi(x)\mu_n^{(\theta)}(dx)}{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n}\right)\right)} \tag{3.17}$$

where

$$\begin{aligned}
\int_{\mathbb{R}}\Phi(x)\mu_n^{(\theta)}(dx) &:= \mathbb{E}\left(\frac{e^{i\theta\gamma_nG}}{\mathbb{E}\left(e^{i\theta\gamma_nG}\right)}\Phi\left(\frac{G}{\gamma_n}\right)\right)\\
&= e^{\theta^2\gamma_n^2/2}\int_{\mathbb{R}} e^{i\theta\gamma_nx}\Phi\left(\frac{x}{\gamma_n}\right)e^{-x^2/2}\frac{dx}{\sqrt{2\pi}}\\
&= \int_{\mathbb{R}}\Phi\left(\frac{x}{\gamma_n}\right)e^{-\frac{1}{2}(x-i\theta\gamma_n)^2}\frac{dx}{\sqrt{2\pi}}\\
&= \int_{\mathbb{R}-i\theta\gamma_n}\Phi\left(\frac{y}{\gamma_n}+i\theta\right)e^{-\frac{1}{2}y^2}\frac{dy}{\sqrt{2\pi}}
\end{aligned}$$

Set

$$g(z) := \Phi(z/\gamma_n + i\theta)e^{-z^2/2}$$

If g is analytic on the whole complex plane, the Cauchy formula gives

$$\int_{[-a,a]} g + \int_{a+i[0,\beta]} g - \int_{[-a,a]+i\beta} g - \int_{-a+i[0,\beta]} g = 0$$

If moreover g satisfies the hypothesis (3.15), we can write

$$\left|\int_{a+i[0,\beta]} g(x)dx\right| \leqslant |\beta|\sup_{z\in a+i[0,\beta]}|g(z)| \xrightarrow[a\to\pm\infty]{} 0$$

Hence,

$$\int_{[-a,a]+i\beta} g = \int_{[-a,a]} g + \left(\int_{a+i[0,\beta]} g - \int_{-a+i[0,\beta]} g \right) =: \int_{[-a,a]} g + R(a)$$

with

$$|R(a)| \leqslant 2\,|\beta| \sup_{z\in a+i[0,\beta]} |g(z)| \xrightarrow[a\to\pm\infty]{} 0$$

Passing to the limit on $a \to +\infty$, we get

$$\int_{\mathbb{R}-i\beta} g = \int_{\mathbb{R}} g$$

Now,

$$\sup_{z\in a+i[0,\beta]} \left| e^{-z^2/2} \right| = \sup_{u\in[0,\beta]} \left| e^{-(a+iu)^2/2} \right| = \sup_{u\in[0,\beta]} e^{-a^2/2+u^2/2} = e^{\beta^2/2} e^{-a^2/2} \xrightarrow[a\to\pm\infty]{} 0$$
$$\sup_{z\in a+i[0,\beta]} |\Phi(z)| = \sup_{u\in[0,\beta]} |\Phi(a+iu)| \xrightarrow[a\to\pm\infty]{} 0 \text{ by the hypothesis (3.15)}$$
$$\sup_{z\in a+i[0,\beta]} \left| e^{-z^2/2} \Phi(z) \right| \xrightarrow[a\to\pm\infty]{} 0$$

We can thus write

$$\int_{\mathbb{R}} \Phi(x) \mu_n^{(\theta)}(dx) = \int_{\mathbb{R}} \Phi\left(\frac{y}{\gamma_n} + i\theta \right) e^{-\frac{1}{2}y^2} \frac{dy}{\sqrt{2\pi}} = \mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n} + i\theta \right) \right)$$

The condition (3.15) ensures that Φ is bounded on a horizontal strip, hence, by the dominated convergence theorem, the continuity of Φ on the complex plane and the hypothesis $\Phi(i\theta) = \Phi(\theta)$ for all $\theta \in \mathbb{R}$, we get

$$\lim_{n\to+\infty} \int_{\mathbb{R}} \Phi(x) \mu_n^{(\theta)}(dx) = \mathbb{E}\left(\lim_{n\to+\infty} \Phi\left(\frac{G}{\gamma_n} + i\theta \right) \right) = \Phi(i\theta) = \Phi(\theta)$$

An by dominated convergence again,

$$\lim_{n\to+\infty} \mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n} \right) \right) = \Phi(0) = 1$$

which proves the theorem.

□

Remark 3.4.9. This theorem is extremely restrictive : most of the examples treated do not satisfy its hypotheses. We will relax several of such hypotheses in forthcoming propositions.

Remark 3.4.10. The fact that the signed (complex) measures $\mu_n^{(\theta)}$ satisfy

$$\lim_{n\to+\infty} \int_{\mathbb{R}} \Phi(x) \mu_n^{(\theta)}(dx) = \Phi(\theta) = \int_{\mathbb{R}} \Phi(x) \delta_\theta(dx)$$

for all Φ satisfying the assumptions of theorem 3.4.7 can be rephrased into a weak convergence of the sequence $(\mu_n^{(\theta)})_n$ to the measure δ_θ. Note that the space of functions on which this convergence holds is restrictive and is a strict subset of the space of continuous bounded functions, space on which the weak convergence does not hold as one can check by considering the limit of the Fourier transform $\int_{\mathbb{R}} e^{i\alpha x}\mu_n^{(\theta)}(dx)$:

$$
\begin{aligned}
\int_{\mathbb{R}} e^{i\alpha x}\mu_n^{(\theta)}(dx) &:= \mathbb{E}\left(\frac{e^{i\theta\gamma_n G}}{\mathbb{E}\left(e^{i\theta\gamma_n G}\right)}e^{i\alpha\frac{G}{\gamma_n}}\right) = \frac{\mathbb{E}\left(e^{iG\left(\theta\gamma_n+\frac{\alpha}{\gamma_n}\right)}\right)}{\mathbb{E}\left(e^{i\theta\gamma_n G}\right)} \\
&= \frac{e^{-\frac{1}{2}\left(\theta\gamma_n+\frac{\alpha}{\gamma_n}\right)^2}}{e^{-\frac{1}{2}(\theta\gamma_n)^2}} = e^{\alpha\theta-\frac{\alpha^2}{2\gamma_n^2}} \xrightarrow[n\to+\infty]{} e^{\alpha\theta}
\end{aligned}
$$

The last theorem motivates the following definition :

Definition 3.4.11. Let Φ be a mod-gaussian limiting function satisfying the hypotheses of theorem 3.4.7. We define the distribution $\mathscr{H}(\Phi,\gamma)$ by

$$
\mathcal{H}_\gamma \sim \mathscr{H}(\Phi,\gamma) \qquad \Longleftrightarrow \qquad \mathbb{P}_{\mathcal{H}_\gamma} := \frac{\Phi\left(\frac{G}{\gamma}\right)}{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma}\right)\right)} \bullet \mathbb{P}_{\gamma G} \tag{3.18}
$$

with $G \sim \mathscr{N}(0,1)$, and with a slight abuse of notation, we also define when the context is clear

$$
\mathcal{H}_n \sim \mathscr{H}(\Phi,\gamma_n) \qquad \Longleftrightarrow \qquad \mathbb{P}_{\mathcal{H}_n} := \frac{\Phi\left(\frac{G}{\gamma_n}\right)}{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n}\right)\right)} \bullet \mathbb{P}_{\gamma_n G} \tag{3.19}
$$

Remark 3.4.12. Let $\mathcal{H}_\gamma \sim \mathscr{H}(\Phi,\gamma)$. Then, for $f \in L^1(\mathbb{P}_G)$

$$
\mathbb{E}\left(f(\mathcal{H}_\gamma)\right) := \frac{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma}\right)f\left(\gamma G\right)\right)}{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma}\right)\right)}
$$

Another way of writing this last formula is to say that $\mathcal{H}_\gamma$ has a Lebesgue-density given by

$$
f_{\mathcal{H}_\gamma}(x) = \frac{1}{c_\gamma}\Phi\left(\frac{x}{\gamma^2}\right)e^{-\frac{1}{2}\left(\frac{x}{\gamma}\right)^2} \quad \text{with} \quad c_\gamma := \gamma\sqrt{2\pi}\mathbb{E}\left(\Phi(G/\gamma)\right) \tag{3.20}
$$

In particular, suppose that a sequence $(X_n)_n$ converges in the mod-gaussian sense to Φ with parameters $(\gamma_n)_n$ and that X_n has a distribution absolutely continuous with respect to the Lebesgue measure, of density f_{X_n}. The Central Limit Theorem amounts to say that

$$
f_{X_n}(x) \text{ is “close to” } \frac{1}{\gamma_n}f_G\left(\frac{x}{\gamma_n}\right)
$$

with $f_G(x) = e^{-x^2/2}/\sqrt{2\pi}$, and the mod-gaussian convergence precises this last fact at the second order since it amounts to say that

$$f_{X_n}(x) \text{ is “close to” } \frac{1}{c_{\gamma_n}} f_G\left(\frac{x}{\gamma_n}\right) \Phi\left(\frac{x}{\gamma_n^2}\right).$$

A formalisation of how close are those last two functions can be achieved by introducing a suitable functionnal distance.

3.4.3 Mod-Gaussian convergence in the Laplace setting

As noticed in remark 3.4.10, the key point in the last theorem is to show that $(\mu_n^{(\theta)})_n$ converges weakly to δ_θ for a certain notion of weak convergence of measures. But the fact that $\lim_{n\to+\infty} \int_{\mathbb{R}} \Phi(x)\mu_n^{(\theta)}(dx) = \Phi(i\theta)$ forces the function Φ to have an additionnal symmetry and gives the hint that this is the variable $i\theta$ that should be the relevant parameter, that is, to consider the Laplace transform in place of the Fourier transform. This motivates the

Definition 3.4.13 (Mod Gaussian-Laplace convergence, [5, 39])**.** Let $(X_n)_n$ be a sequence of random variables of expectation 0 with $\mathbb{E}\left(e^{uX_n}\right) < \infty$ for all $u \in \mathbb{R}$ and let $(\gamma_n)_n$ be a sequence of strictly positive real numbers. $(X_n)_n$ is said to converge mod-Gaussian-Laplace with parameters $(\gamma_n)_n$ if

$$\frac{\mathbb{E}\left(e^{uX_n}\right)}{\mathbb{E}\left(e^{u\gamma_n G}\right)} \xrightarrow[n\to+\infty]{} \Phi(u)$$

where $\Phi : \mathbb{R} \to \mathbb{R}_+$ is a continuous function satisfying $\Phi(0) = 1$, the last convergence being locally uniform in $u \in \mathbb{R}$.

Note that the function Φ here defined is positive, as a limit of a sequence of positive functions. Now, let $(\mathbb{P}_{\xi_n(u)})_n$ be the probability measure defined by

$$\mathbb{P}_{\xi_n(u)} := \frac{e^{u\gamma_n G}}{\mathbb{E}\left(e^{u\gamma_n G}\right)} \bullet \mathbb{P}_{G/\gamma_n} = \mathbb{P}_{G/\gamma_n+u}$$

by the natural generalisation of (3.13) to variances different of 1. We hence have

$$\mathbb{E}\left(e^{i\theta\xi_n(u)}\right) = \mathbb{E}\left(e^{i\theta\left(\frac{G}{\gamma_n}+u\right)}\right) = e^{i\theta u - \frac{\theta^2}{2\gamma_n^2}} \xrightarrow[n\to+\infty]{} e^{i\theta u}$$

that is :

$$\xi_n(u) \stackrel{\mathcal{L}}{=} \frac{G}{\gamma_n} + u \xrightarrow[n\to+\infty]{\mathcal{L}} u$$

Suppose that $\Phi \in L^1\left(\mathbb{P}_{G/\gamma_n}\right)$ for all $n \geqslant 0$. We would like to have the equality

$$\frac{\mathbb{E}\left(e^{-x\mathcal{H}_{\gamma_n}(\Phi)}\right)}{\mathbb{E}\left(e^{-x\gamma_n G}\right)} = \frac{\mathbb{E}\left(\Phi\left(x + \frac{G}{\gamma_n}\right)\right)}{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma_n}\right)\right)} \tag{3.21}$$

in the same vein as (3.17) but for a limiting function in the Laplace sense ; such an equality can be understood as a duality equality between mod-Gaussian convergence and convergence in law since the convergence of the LHS is the definition of mod-Gaussian convergence and the RHS converges because of the convergence in law of G/γ_n for test functions being bounded and continuous (note that we did not suppose Φ bounded). As Φ is positive, the last interpretation by means of a penalisation holds. But to pass to the limit, one needs the boundedness of Φ to apply dominated convergence.

Supposing now that Φ is bounded, one can define the probability distribution $\widetilde{\mathscr{H}}(\Phi, \gamma)$ by

$$\widetilde{\mathcal{H}}_\gamma \sim \widetilde{\mathscr{H}}(\Phi, \gamma) \iff \mathbb{P}_{\widetilde{\mathcal{H}}_\gamma} := \frac{\Phi\left(\frac{G}{\gamma}\right)}{\mathbb{E}\left(\Phi\left(\frac{G}{\gamma}\right)\right)} \bullet \mathbb{P}_{\gamma G}$$

As (3.21) holds, an application of dominated convergence and continuity of Φ gives

$$\frac{\mathbb{E}\left(e^{-x\mathcal{H}_{\gamma_n}(\Phi)}\right)}{\mathbb{E}\left(e^{-x\gamma_n G}\right)} \xrightarrow[n\to+\infty]{} \Phi(x)$$

locally uniformly in x. In particular, for $(X_n)_n$ converging in the mod-Gaussian-Laplace sense with parameters $(\gamma_n)_n$, we have

$$\frac{\mathbb{E}\left(e^{x\widetilde{\mathcal{H}}_{\gamma_n}}\right)}{\mathbb{E}\left(e^{xX_n}\right)} \xrightarrow[n\to+\infty]{} 1 \quad \text{if } \Phi(x) \neq 0$$

In accordance with theorem 3.4.7, the distribution of $(X_n)_n$ should look like the distribution of $(\widetilde{\mathcal{H}}_{\gamma_n})_n$. To precise this idea of resemblance of the distributions, the best way is to introduce a probabilistic metric, such as the Kolmogorov one, and to compute the effective distance between $(X_n)_n$ and $(\widetilde{\mathcal{H}}_{\gamma_n})_n$. This is the object of the chapter 4.7.

Of course, definition 3.4.13 can be adapted in case $\mathbb{E}\left(e^{uX_n}\right) < \infty$ only on a strict subset of $\mathbb{R}$.

3.4.4 The modulus of the characteristic polynomial of a random unitary matrix

Barnes' asymptotics for the G-function gives (see [8])

$$\log G(1+z) = \frac{1}{12}\log A + \log(2\pi)\frac{z}{2} + \left(\frac{z^2}{2} - \frac{1}{12}\right)\log z - \frac{3}{4}z^2 + O\left(\frac{1}{z^2}\right)$$

valid for z in any sector not containing the negative real axis with $|z|$ large. Thus, for $x > 0$, one has

$$\widetilde{\Phi}_U(x) = \frac{G(1+x)^2}{G(1+2x)} = e^{-x^2\log|x| - O(x^2)}$$

i.e. we are in the case where the assumptions of definition 3.4.13 apply. Unfortunately, in the neighborhood of $x = -1/2$, $\widetilde{\Phi}_U$ has the following type of singularity

$$\widetilde{\Phi}_U(x) \underset{x\to -1/2}{\sim} \frac{A}{x+1/2}$$

and the function is not integrable in the neighbourhood of $-1/2$. The functions $\widetilde{\Phi}_U$ or Φ_U are hence only integrable (for the Gaussian measure) on $[-1/2+\varepsilon, +\infty[$ for all $\varepsilon > 0$.

A possible way to define the mod-Gaussian-Laplace convergence on $]-1/2, +\infty[$ consists in defining first the random variables $H_n(\Phi_\varepsilon) \sim \widetilde{\mathscr{H}}(\Phi_\varepsilon, \gamma)$ using $\Phi_\varepsilon(x) := \Phi_U(x)\mathbb{1}_{\{x>-1/2+\varepsilon\}}$ in place of Φ_U. By continuity and dominated convergence, locally uniformly in $x > -1/2$,

$$\frac{\mathbb{E}\left(e^{-xH_n(\Phi_\varepsilon)}\right)}{\mathbb{E}\left(e^{-x\gamma_n G}\right)} \xrightarrow[n\to+\infty]{} \Phi_\varepsilon(x)$$

To get the whole interval $]-1/2, +\infty[$, a *diagonal extraction procedure* can be applied to get a sequence $(\varepsilon_n)_n$. The sequence $(H_n(\Phi_{\varepsilon_n}))_n$ will then converge in the mod-Gaussian-Laplace sense to $\Phi_0 : x \mapsto \Phi_U(x)\mathbb{1}_{\{x>-1/2\}}$ which is the desired function (i.e. on the good interval of definition).

In the general case, this truncation procedure has to be defined everytime the limiting mod-* function is not integrable for the relevant mode of convergence. This is for instance the case of $\log\left|\zeta\left(\frac{1}{2}+iTU\right)\right|$: to give a probabilistic flavour of the moments conjecture, one has to define the suitable sequence of truncated random variables. Note that Selberg's CLT can be rephrased into

$$d_{\mathrm{Kol}}\left(\frac{\log\left|\zeta\left(\frac{1}{2}+iTU\right)\right|}{\sqrt{\frac{1}{2}\log\log T}}, \mathscr{N}(0,1)\right) \xrightarrow[T\to+\infty]{} 0$$

Nevertheless, a translation of the moments conjecture into

$$d_{\mathrm{Kol}}\left(\log\left|\zeta\left(\frac{1}{2}+iTU\right)\right|, \widetilde{\mathscr{H}}(\Phi_{U,\varepsilon_T}\Phi_{A,\varepsilon'_T}, \log\log T)\right) \xrightarrow[T\to+\infty]{} 0$$

is not sufficient : one needs to include functions such as $x \mapsto e^{\lambda x}$ into the space of test functions defining the distance to be able to renormalise by the Laplace transform of $\log\log TG$.

Remark 3.4.14. The choosen truncation could also be done in a different maneer. Indeed, since every limiting mod-* function can be written as a product (appearing on the examples considered, and abstract in the case of a general sequence of random variables), one can look for a truncation of the number of terms of the product, if nevertheless this product appears to be integrable for a certain tuncation.

3.4.5 Mod-Poisson convergence in the Laplace setting

In the same vein as before, we define the mod-Poisson convergence in the Laplace setting by the following :

Definition 3.4.15 (Mod Poisson-Laplace convergence, [5, 39])**.** Let $(X_n)_n$ be a sequence of positive random variables and $(\gamma_n)_n$ be a sequence of strictly positive real numbers. $(X_n)_n$ is said to converge mod-Poisson-Laplace with parameters $(\gamma_n)_n$ if for all $x \in]0, +\infty[$

$$\frac{\mathbb{E}\left(x^{X_n}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} \xrightarrow[n\to+\infty]{} \Phi(x)$$

where $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous function satisfying $\Phi(1) = 1$, the last convergence being locally uniform in $x \in]0, +\infty[$, and $P_{\gamma_n} \sim \mathscr{P}(\gamma_n)$.

Note that the function Φ is necessarily positive on $\mathbb{R}_+^*$ as a locally uniform limit of positive functions on $\mathbb{R}_+^*$. If $\mathbb{E}\left(\Phi(P_\gamma/\gamma)\right) < \infty$ for all $\gamma \in \mathbb{R}_+$, this positivity allows to define the probability distribution $\mathscr{Q}\left(\Phi, \gamma\right)$ by

$$Q_\gamma \sim \mathscr{Q}\left(\Phi, \gamma\right) \iff \mathbb{P}_{Q_\gamma} := \frac{\Phi\left(\frac{P_\gamma}{\gamma}\right)}{\mathbb{E}\left(\Phi\left(\frac{P_\gamma}{\gamma}\right)\right)} \bullet \mathbb{P}_{P_\gamma}$$

where $P_\gamma \sim \mathscr{P}(\gamma)$. For $Q_{\gamma_n} \sim \mathscr{Q}\left(\Phi, \gamma_n\right)$, we have

$$\frac{\mathbb{E}\left(x^{Q_{\gamma_n}}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} = \frac{\mathbb{E}\left(\Phi\left(\frac{P_{\gamma_n}}{\gamma_n}\right) x^{P_{\gamma_n}}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)\mathbb{E}\left(\Phi\left(\frac{P_{\gamma_n}}{\gamma_n}\right)\right)} = \frac{\mathbb{E}\left(\frac{x^{P_{\gamma_n}}}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)}\Phi\left(\frac{P_{\gamma_n}}{\gamma_n}\right)\right)}{\mathbb{E}\left(\Phi\left(\frac{P_{\gamma_n}}{\gamma_n}\right)\right)} =: \frac{\mathbb{E}\left(\Phi\left(\pi_n(x)\right)\right)}{\mathbb{E}\left(\Phi\left(\frac{P_{\gamma_n}}{\gamma_n}\right)\right)}$$

with

$$\mathbb{P}_{\pi_n(x)} := \frac{x^{P_{\gamma_n}}}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} \bullet \mathbb{P}_{P_{\gamma_n}/\gamma_n} = \mathbb{P}_{P_{x\gamma_n}/\gamma_n}$$

since

$$\mathbb{E}\left(e^{i\theta\pi_n(x)}\right) = \frac{\mathbb{E}\left(e^{i\frac{\theta}{\gamma_n}P_{\gamma_n}} x^{P_{\gamma_n}}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} = \frac{e^{\gamma_n\left(xe^{i\theta/\gamma_n}-1\right)}}{e^{\gamma_n(x-1)}} = e^{\gamma_n x\left(e^{i\theta/\gamma_n}-1\right)} = \mathbb{E}\left(e^{i\frac{\theta}{\gamma_n}P_{x\gamma_n}}\right)$$

that is :

$$\pi_n(x) \overset{\mathcal{L}}{=} \frac{P_{x\gamma_n}}{\gamma_n} \xrightarrow[n\to+\infty]{\mathcal{L}} x$$

Hence, $(Q_{\gamma_n})_n$ converges in the mod-Poisson-Laplace sense to Φ, and $\mathbb{E}\left(x^{Q_{\gamma_n}}\right)/\mathbb{E}\left(x^{X_n}\right)$ converges to 1 for every $x \in \mathbb{R}_+^*$ such that $\Phi(x) \neq 0$, which expresses that X_n "looks like" Q_{γ_n}. A more quantitative version of this result could be given by means of Stein's method by computing $d_{\mathrm{Kol}}\left(X_n, Q_{\gamma_n}\right)$.

Remark 3.4.16. Note the similarity between the duality formula (3.21) and

$$\frac{\mathbb{E}\left(x^{\mathcal{Q}_\gamma(\Phi)}\right)}{\mathbb{E}\left(x^{P_\gamma}\right)} = \frac{\mathbb{E}\left(\Phi\left(\frac{P_{x\gamma}}{\gamma}\right)\right)}{\mathbb{E}\left(\Phi\left(\frac{P_\gamma}{\gamma}\right)\right)} \tag{3.22}$$

There is a formal correspondance between $\mathscr{P}(x\gamma)$ and $\mathscr{N}(x\gamma, 1)$ since one can write (3.21) with $N_\gamma \sim \mathscr{N}(\gamma, 1)$ as

$$\frac{\mathbb{E}\left(e^{-x\mathcal{H}_\gamma(\Phi)}\right)}{\mathbb{E}\left(e^{-xN_\gamma}\right)} = \frac{\mathbb{E}\left(\Phi\left(\frac{N_{x\gamma}}{\gamma}\right)\right)}{\mathbb{E}\left(\Phi\left(\frac{N_\gamma}{\gamma}\right)\right)}$$

This comes from the particular type of change of probability of $\mathscr{P}(\gamma)$ and $\mathscr{N}(\gamma, 1)$. In the case of a more general Lévy process, e.g. a subordinator $(X_\gamma)_\gamma$ of drift $\bar{\mathrm{d}}$ and Lévy measure Π, the Lévy-Kintchine formula gives

$$\mathbb{E}\left(e^{-\theta X_\gamma}\right) = \exp\left(-\gamma \Lambda_X(\theta)\right)$$

with

$$\Lambda_X(\theta) = \bar{\mathrm{d}}\theta + \int_0^{+\infty} \left(1 - e^{-\theta u}\right) \Pi(du)$$

Using the Lévy-Kintchine formula, one has

$$\frac{e^{-xX_\gamma}}{\mathbb{E}\left(e^{-xX_\gamma}\right)} \bullet \mathbb{P}_{X_\gamma} =: \mathbb{P}_{X_\gamma^{(x)}}$$

where $(X_\gamma^{(x)})_\gamma$ is the subordinator of drift $\bar{\mathrm{d}}$ and Lévy measure $\Pi^{(x)}$ given by the exponential bias

$$\Pi^{(x)}(du) = e^{-xu}\Pi(du)$$

Thus, if one defines $\mathscr{X}(\Phi, \gamma)$ (under the restriction of its existence) by

$$\mathcal{X}_\gamma(\Phi) \sim \mathscr{X}(\Phi, \gamma) \iff \mathbb{P}_{\mathcal{X}_\gamma(\Phi)} := \frac{\Phi\left(\frac{X_\gamma}{\gamma}\right)}{\mathbb{E}\left(\Phi\left(\frac{X_\gamma}{\gamma}\right)\right)} \bullet \mathbb{P}_{X_\gamma}$$

one has

$$\frac{\mathbb{E}\left(e^{-x\mathcal{X}_\gamma(\Phi)}\right)}{\mathbb{E}\left(e^{-xX_\gamma}\right)} = \frac{\mathbb{E}\left(\Phi\left(\frac{X_\gamma^{(x)}}{\gamma}\right)\right)}{\mathbb{E}\left(\Phi\left(\frac{X_\gamma}{\gamma}\right)\right)}$$

But, using the Lévy-Kintchine formula again,

$$\frac{X_\gamma^{(x)}}{\gamma} \xrightarrow[\gamma \to +\infty]{\mathcal{L}} \bar{\mathrm{d}} + \int_0^{+\infty} u\,\Pi^{(x)}(du) = \bar{\mathrm{d}} + \int_0^{+\infty} u\,e^{-ux}\,\Pi(du) = \Lambda_X'(x)$$

hence, if Φ is continuous and bounded, by dominated convergence,

$$\frac{\mathbb{E}\left(e^{-x\mathcal{X}_\gamma(\Phi)}\right)}{\mathbb{E}\left(e^{-xX_\gamma}\right)} \xrightarrow[\gamma\to+\infty]{} \frac{\Phi\left(\Lambda_X'(x)\right)}{\Phi\left(\Lambda_X'(0)\right)}$$

Set

$$\Upsilon(x) := \inf\left\{y \in \mathbb{R} \;/\; \Lambda_X'(x) \geqslant y\right\}$$

Supposing this last quantity finite, we have

$$\frac{\mathbb{E}\left(e^{-\Upsilon(x)\,\mathcal{X}_\gamma(\Phi)}\right)}{\mathbb{E}\left(e^{-\Upsilon(x)\,X_\gamma}\right)} \xrightarrow[\gamma\to+\infty]{} \frac{\Phi\left(x\right)}{\Phi\left(1\right)}$$

For example, in the case of the Poisson process, since $\Pi = \delta_1$, we have $\Upsilon(x) = -\log x$, which gives (3.22). The case of the stable subordinator (without the Cauchy process) can be handled in the same way with $\Upsilon(x) = x^{1/(\alpha-1)}$ since, for $x > 0$, $\alpha \neq 1$ and $\bar{\mathrm{d}} = 0$

$$\frac{X_\gamma^{(x)}}{\gamma} \xrightarrow[\gamma\to+\infty]{\mathcal{L}} \alpha x^{\alpha-1}$$

Example 3.4.17. Let $D \sim \mathscr{D}(1)$ be a random variable Dickman-distributed. The Dickman distribution is defined by the equality (see e.g. [2])

$$\mathbb{E}\left(e^{i\theta D}\right) = \exp\left(\int_0^1 \left(e^{i\theta u} - 1\right)\frac{du}{u}\right)$$

The Lévy measure is here

$$\Pi_D(du) = \mathbb{1}_{\{0\leqslant u\leqslant 1\}}\frac{du}{u}$$

and one has

$$\Lambda_D'(x) = \int_0^{+\infty} u e^{-xu}\Pi_D(du) = \int_0^1 e^{-xu}du = \frac{1-e^{-x}}{x}$$

Since Λ_D' is strictly decreasing on $\mathbb{R}_+$, it is a bijection from $\mathbb{R}_+$ to $(0,1]$ and the inverse bijection Υ can be defined. In order to apply the precedent construction, we need some sequences of random variables that converge in law to a Dickman distribution. This distribution arises, amongst others, in several models of the form

$$D_n := \sum_{k=1}^n kZ_k$$

with $(Z_k)_k$ a sequence of weakly correlated random variables with values in $\mathbb{N}$ that have a certain probability of being equal to 0, typically $\mathbb{P}(Z_k = 0) = 1/k$. For instance, if Z_k is Bernoulli, Poisson or Geometrically distributed with parameter $1/k$, the last sum D_n/n will converge in law to the Dickman distribution, as one can see with $Z_k \sim \mathscr{P}(1/k)$, the Z_k being independent

$$\begin{aligned}\mathbb{E}\left(e^{i\theta D_n/n}\right) &= \prod_{k=1}^n \mathbb{E}\left(e^{i\theta Z_k/n}\right) = \prod_{k=1}^n \exp\left(\frac{e^{i\theta k/n}-1}{k}\right) = \exp\left(\frac{1}{n}\sum_{k=1}^n \frac{e^{i\theta k/n}-1}{k/n}\right)\\ &\xrightarrow[n\to+\infty]{} \exp\left(\int_0^1 \left(e^{i\theta u}-1\right)\frac{du}{u}\right)\end{aligned}$$

The mod-Dickman convergence can be defined in the same way as before. In particular, one has in the Bernoulli case, for $D_n = \sum_{1\leqslant k\leqslant n} kB_k$ and $x \geqslant 0$

$$\begin{aligned}\frac{\mathbb{E}\left(e^{-xD_n}\right)}{\mathbb{E}\left(e^{-xnD}\right)} &= e^{-\int_0^n\left(e^{-xu}-1\right)\frac{du}{u}}\prod_{k=1}^n \mathbb{E}\left(e^{-xB_k}\right) = \prod_{k=1}^n\left(1+\frac{e^{-kxu}-1}{k}\right)e^{-\int_{k-1}^k\left(e^{-xu}-1\right)\frac{du}{u}}\\ &\xrightarrow[n\to+\infty]{} \Phi_B(x) := \prod_{k\geqslant 1}\left(1+\frac{e^{-kxu}-1}{k}\right)e^{-\int_{k-1}^k\left(e^{-xu}-1\right)\frac{du}{u}}\end{aligned}$$

the last product being convergent, as one can check by developing $\log\left(1+\frac{e^{-xu}-1}{k}\right)$ and using the rest of the Riemann sum approximation.

In the independent Poisson case, with $P_k \sim \mathscr{P}(1/k)$, one has

$$\frac{\mathbb{E}\left(e^{-x\widetilde{D}_n}\right)}{\mathbb{E}\left(e^{-xnD}\right)} = e^{-\int_0^n \left(e^{-xu}-1\right)\frac{du}{u}} \prod_{k=1}^n \mathbb{E}\left(e^{-xP_k}\right) = \exp\left(\sum_{k=1}^n \left(\frac{e^{-kxu}-1}{k} - \int_{k-1}^k \left(e^{-xu}-1\right)\frac{du}{u}\right)\right)$$
$$\xrightarrow[n\to+\infty]{} \Phi_P(x) := \exp\left(\sum_{k\geqslant 1}\left(\frac{e^{-kxu}-1}{k} - \int_{k-1}^k \left(e^{-xu}-1\right)\frac{du}{u}\right)\right)$$

the last sum being convergent for the same reason as in the Bernoulli case, using the error term of a Riemann sum approximation.

The case of independent geometric random variables can also be treated, but with a sum starting at $k = 2$. The limiting function has then a slight corrective term :

$$\begin{aligned}\frac{\mathbb{E}\left(e^{-x\widehat{D}_n}\right)}{\mathbb{E}\left(e^{-xnD}\right)} &= e^{-\int_0^n \left(e^{-xu}-1\right)\frac{du}{u}} \prod_{k=2}^n \mathbb{E}\left(e^{-xkG_k}\right) = e^{-\int_0^n \left(e^{-xu}-1\right)} \prod_{k=2}^n \left(\frac{1-\frac{1}{k}}{1-\frac{e^{-kx}}{k}}\right) \\ &= e^{-\int_0^1 \left(e^{-xu}-1\right)\frac{du}{u}} \frac{1}{n} \prod_{k=2}^n \frac{e^{-\int_{k-1}^k \left(e^{-xu}-1\right)\frac{du}{u}}}{1-\frac{e^{-kx}}{k}} \\ &\xrightarrow[n\to+\infty]{} \Phi_G(x)\end{aligned}$$

with

$$\begin{aligned}\Phi_G(x) &= \exp\left(-\gamma - \int_0^1 \left(e^{-xu}-1\right)\frac{du}{u} - \sum_{k\geqslant 2}\left(\log\left(1-\frac{e^{-kx}}{k}\right) + \frac{1}{k} + \int_{k-1}^k \left(e^{-xu}-1\right)\frac{du}{u}\right)\right) \\ &= \exp\left(-\int_0^1 \left(e^{-xu}-1\right)\frac{du}{u} - \sum_{k\geqslant 2}\left(\log\left(1-\frac{e^{-kx}}{k}\right) + \int_{k-1}^k e^{-xu}\frac{du}{u}\right)\right)\end{aligned}$$

In these three cases, the limiting function $\Phi \in \{\Phi_B, \Phi_P, \Phi_G\}$ is positive and bounded by 1 ; hence, the associated random variables $\mathcal{X}_\gamma(\Phi)$ can be defined (with $\gamma_n = n$).

Remark 3.4.18. In the last three examples, the only case to understand concerns a Bernoulli random variable B_γ with $0 < \gamma < 1$: one can easily check that there exist random variables $Z_1(\gamma)$ and $Z_2(\gamma)$ independent of $P_\gamma \sim \mathscr{P}(\gamma)$ and $G_\gamma \sim \mathrm{Ge}(\gamma)$ such that (see [70] for the Poisson case, the generalisation to the Geometric case is straightforward)

$$\begin{aligned}P_\gamma &\stackrel{\mathcal{L}}{=} B_\gamma + Z_1(\gamma) \\ G_\gamma &\stackrel{\mathcal{L}}{=} B_\gamma + Z_2(\gamma)\end{aligned}$$

3.5 Constructing mod-* fluctuations

3.5.1 Comparing models

Back to the Rényi-Túran proof (c.f. [83]) of the Erdös-Kac theorem : for $U_n \sim \mathscr{U}(\llbracket 1, n \rrbracket)$, recall that $\omega(U_n)$ indicates the number of prime divisors of U_n, and that $(\omega(U_n), H_n^{(\mathcal{P})})_n$ converges in the mod-Poisson sense to

$$\begin{aligned}\Phi_\omega(z) &= \Phi_C(z)\Phi_\Omega(z) = \prod_{k\in\mathbb{N}^*}\left(1+\frac{z-1}{k}\right)e^{-\frac{z-1}{k}}\prod_{p\in\mathcal{P}}\left(1+\frac{z-1}{p}\right)e^{-\frac{z-1}{p}}\\ &= \prod_{k\in\mathbb{N}^*}\left[\left(1+\frac{z-1}{k}\right)e^{-\frac{z-1}{k}}\right]^{1+\mathbb{1}_{\{k\in\mathcal{P}\}}}\end{aligned}$$

defined in (3.11), with $H_n^{(\mathcal{P})}$ defined in (3.8). It was proven that for all $x \geqslant 0$, $\Phi_\omega(x) \geqslant 0$, hence, it only remains to check that $\Phi_\omega(\cdot/\gamma) \in \ell^1(\mathscr{P}(\gamma))$ for $\gamma = H_n^{(\mathcal{P})}$. But for all $y \in \mathbb{R}$, one has $1 + y \leqslant e^y$, hence, for $y = \frac{x-1}{k}$, one has for all $k \geqslant 1$ and for all $x \in \mathbb{R}$

$$\left(1+\frac{x-1}{k}\right)e^{-\frac{x-1}{k}} \leqslant 1$$

which implies that for all $x \in \mathbb{R}$ (hence for all $x \in \mathbb{R}_+$)

$$\Phi_\omega(x) \leqslant 1$$

Finally, we can conclude that $\omega(U_n)$ has a distribution closed to $\mathscr{Q}\left(\Phi_\omega, H_n^{(\mathcal{P})}\right)$ and one can expect $d_{\mathrm{Kol}}\left(\omega(U_n), \mathscr{Q}\left(\Phi_\omega, H_n^{(\mathcal{P})}\right)\right)$ to tend to 0 when $n \to +\infty$; this result can be achieved by means of Stein's method (see chapter 4).

One can ask the question of a pathwise construction of such a biased random variable in the same vein as lemma 3.4.5, or more generally, to exhibit a simple random variable that converges in the mod-Poisson sense to Φ_ω. Such a random variable would thus be, somehow, a more accurate model than the usual independent sum as it would converge at the second order. Of course, this assertion must be precised : one intuitive way to construct such a model can be done with a sum of independent sums of random variables created by means of Bernoulli random variables, i.e.

$$\omega(U_n) \approx \omega_n := \sum_{k=1}^{A} B_k + \sum_{k=1}^{A'} B'_k \tag{3.23}$$

with $\mathbb{P}(B_k = 1) = \frac{1}{\mathrm{p}_k}$, $\mathbb{P}(B'_k = 1) = \frac{1}{k}$, where we have set $\mathcal{P} := \{\mathrm{p}_k, k \geqslant 1\}$ and where A, A' are choosen so that the mod-Poisson speed of convergence $\gamma_n = \log\log n + \kappa$ of $\omega(U_n)$ matches the speed of convergence of ω_n. Since γ_n is asymptotically the mean (and the variance) of $\omega(U_n)$, and since

$$\mathbb{E}(\omega_n) = \sum_{k=1}^{A}\frac{1}{\mathrm{p}_k} + \sum_{k=1}^{A'}\frac{1}{k} = \log\log A + \log(A') + O(1)$$

one finds the relation

$$A' \log A = O(\log n)$$

the constant in the O being explicitely known.

This intuitive model, despite its artificial character, has the advantage of being an acceptable mod-Poisson model for $\omega(U_n)$ since it converges in the mod-Poisson sense to $\Phi_\omega = \Phi_\Omega \Phi_C$, but it gives no hint about a number-theoretic interpretation of the apparition of this term, as it is done in [68] in the framework of function fields.

Nevertheless, the question of constructing such a model can be asked, the relevance of such a model still being to be debated.

3.5.2 Construction of the model

Theorem 3.5.1 (A mod-Poisson model of the Erdös-Kac theorem, [6])**.** *Let $(B_k)_k$ and $(B'_k)_k$ be two independent sequences of independent $\{0,1\}$-Bernoulli random variables such that*

$$\mathbb{P}\left(B_k = 1\right) = \mathbb{P}\left(B'_k = 1\right) = \frac{1}{k}$$

For $\theta > 0$, let $B_k(\theta)$ be the θ-exponential bias of $B_k = B_k(1)$ given by

$$\mathbb{P}\left(B_k(\theta) = 1\right) = \frac{\theta}{\theta + k - 1} = 1 - \mathbb{P}\left(B_k(\theta) = 0\right)$$

or equivalently

$$\mathbb{P}_{B_k(\theta)} := \frac{\theta^{B_k}}{\mathbb{E}\left(\theta^{B_k}\right)} \bullet \mathbb{P}_{B_k}$$

Let

$$\begin{aligned}
\gamma_n &:= H_n^{(\mathcal{P})} \underset{n\to+\infty}{\sim} \log\log n \\
k_n &:= [\sqrt{\gamma_n}] \underset{n\to+\infty}{\sim} \sqrt{\log\log n} \\
\pi(n) &:= \sum_{p\in\mathcal{P}} \mathbb{1}_{\{p\leqslant n\}} \underset{n\to+\infty}{\sim} \frac{n}{\log n} \\
v_n &:= \exp\left(-\frac{H_{k_n}}{\gamma_n}\right) = \exp\left(-\frac{\log\log\log n + O(1)}{2\log\log n}\right)
\end{aligned}$$

and

$$C'_n := \sum_{\ell=1}^{k_n} B'_\ell(1/\gamma_n)$$

Let $(I_\ell)_\ell$ be a sequence of i.i.d. random variables in $[\![1, \pi(n)]\!]$ independent of $(B_k)_k$ and $(B'_k)_k$ distributed according to

$$\mathbb{P}\left(I = k\right) = \frac{\frac{1}{\mathrm{p}_k + v_n + 1}}{\sum_{\ell=1}^{\pi(n)} \frac{1}{\mathrm{p}_\ell + v_n + 1}}$$

for all $k \in [\![1, \pi(n)]\!]$. Let $\delta(I_1, \dots, I_k)$ be the length of the random partition created by means of the paintbox process associated to $(I_\ell)_\ell$. Then, the random variable Ω''_n defined by

$$\Omega''_n := \sum_{k \neq I_1, \dots, I_{C'_n}} B_{\mathrm{p}_k}(v_n) + \delta(I_1, \dots, I_{C'_n})$$

is such that

$$\left(\Omega''_n, H_n^{(\mathcal{P})}\right) \xrightarrow[n \to +\infty]{\text{mod-P}} \Phi_\omega$$

As a consequence of mod-Poisson convergence, one has in addition the CLT

$$\frac{\Omega''_n - \log\log n}{\sqrt{\log\log n}} \xrightarrow[n \to +\infty]{\mathcal{L}} \mathscr{N}(0, 1)$$

Proof. The idea to construct such a random variable lies on two approximations : approximate the random variable $P_{\gamma_n} \sim \mathscr{P}(\gamma_n)$ with the independent model Ω_n since at the first order of convergence (i.e. convergence in law) these random variables are close, and approximate the limiting function Φ_ω by a suitable truncation $\Phi_\omega^{(n)}$ of its product since such a finite product converges locally uniformly to Φ_ω.

First step : Change of random variable : Set $\gamma_n := H_n^{(\mathcal{P})}$ and

$$\begin{aligned}
\mathbb{P}_{\Omega'_n} &:= \frac{\Phi_C\left(\frac{\Omega_n}{\gamma_n}\right)}{\mathbb{E}\left(\Phi_C\left(\frac{\Omega_n}{\gamma_n}\right)\right)} \bullet \mathbb{P}_{\Omega_n} \\
\mathbb{P}_{\Omega_n(x)} &:= \frac{x^{\Omega_n}}{\mathbb{E}\left(x^{\Omega_n}\right)} \bullet \mathbb{P}_{\Omega_n}
\end{aligned}$$

and remark that

$$\frac{\mathbb{E}\left(x^{\Omega'_n}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} = \frac{\mathbb{E}\left(\Phi_C\left(\frac{\Omega_n}{\gamma_n}\right) x^{\Omega_n}\right)}{\mathbb{E}\left(\Phi_C\left(\frac{\Omega_n}{\gamma_n}\right)\right) \mathbb{E}\left(x^{\Omega_n}\right)} \frac{\mathbb{E}\left(x^{\Omega_n}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} = \frac{\mathbb{E}\left(\Phi_C\left(\frac{\Omega_n(x)}{\gamma_n}\right)\right)}{\mathbb{E}\left(\Phi_C\left(\frac{\Omega_n}{\gamma_n}\right)\right)} \frac{\mathbb{E}\left(x^{\Omega_n}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)}$$

By mod-Poisson convergence, we have, locally uniformly in $x \in \mathbb{R}_+$

$$\frac{\mathbb{E}\left(x^{\Omega_n}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} \xrightarrow[n \to +\infty]{} \Phi_\Omega(x)$$

By dominated convergence, continuity of Φ_C and law of large numbers for Ω_n that implies that $\Omega_n / \gamma_n \to 1$ almost surely and in L^1, we have

$$\mathbb{E}\left(\Phi_C\left(\frac{\Omega_n}{\gamma_n}\right)\right) \xrightarrow[n \to +\infty]{} \Phi_C(1) = 1$$

Last, using lemma 3.4.6, we see that

$$\Omega_n(x) \stackrel{\mathcal{L}}{=} \sum_{p \leqslant n, p \in \mathcal{P}} B_p(x)$$

with $\mathbb{P}\left(B_p(x)=1\right)=\frac{x/p}{x/p+1-1/p}=\frac{x}{p+x-1}$. Using the law of large numbers and the fact that

$$\frac{\sum_{p\leqslant n}\frac{x}{p+x-1}}{\sum_{p\leqslant n}\frac{1}{p}}\xrightarrow[n\to+\infty]{}x$$

we get that $\Omega_n(x)/\gamma_n\to x$ almost surely[5] and in L^1, which implies by continuity of Φ_C and dominated convergence that

$$\mathbb{E}\left(\Phi_C\left(\frac{\Omega_n(x)}{\gamma_n}\right)\right)\xrightarrow[n\to+\infty]{}\Phi_C(x)$$

Hence, we have a first random variable that converges in the mod-Poisson sense to $\Phi_\omega=\Phi_C\Phi_\Omega$:

$$\left(\Omega_n', H_n^{(\mathcal{P})}\right)\xrightarrow[n\to+\infty]{\text{mod-P}}\Phi_\omega$$

<u>**Second step :**</u> Truncation of Φ_C : Let $k\in\mathbb{N}^*$ and

$$\Phi_C^{(k)}(x):=\prod_{\ell=1}^{k}\left(1+\frac{x-1}{\ell}\right)e^{-\frac{x-1}{\ell}}$$

We clearly have for all $x\in\mathbb{R}$

$$\Phi_C^{(k)}(x)\leqslant 1$$

since every term of the product is less than 1. Elementary computations show that

$$\begin{aligned}\left\|\Phi_C^{(k)}-\Phi_C\right\|_\infty &= O\left(\frac{1}{k}\right)\\ \left\|D\Phi_C^{(k)}\right\|_\infty &= O\left(1\right)\end{aligned}$$

where the supremum is taken for $x\in\mathbb{R}_+$ and where $Df(x):=f'(x)$. Last,

$$\left|\mathbb{E}\left(\Phi_C^{(k)}\left(\frac{\Omega_n(x)}{\gamma_n}\right)\right)-\Phi_C(x)\right|\leqslant\left\|\Phi_C^{(k)}-\Phi_C\right\|_\infty+\left\|D\Phi_C^{(k)}\right\|_\infty\mathbb{E}\left(\left|\frac{\Omega_n(x)}{\gamma_n}-x\right|\right)$$

The classical CLT for sums of independent random variables ensures that

$$\mathbb{E}\left(\left|\frac{\Omega_n(x)}{\gamma_n}-x\right|\right)=O\left(\frac{1}{\sqrt{\gamma_n}}\right)$$

with an absolute constant in the $O(\cdot)$. Thus, taking

$$k=k_n:=\sqrt{\gamma_n}$$

one has

$$\left|\mathbb{E}\left(\Phi_C^{(k)}\left(\frac{\Omega_n(x)}{\gamma_n}\right)\right)-\Phi_C(x)\right|=O\left(\frac{1}{\sqrt{\gamma_n}}\right)$$

[5] We suppose that x and the Bernoulli random variables are defined on the same probability space.

With this choice of k_n, one has moreover

$$v_n := \exp\left(-\frac{H_{k_n}}{\gamma_n}\right) = \exp\left(-\frac{\log\log\log n + O(1)}{2\log\log n}\right)$$

We can now define

$$\mathbb{P}_{\Omega''_n} := \frac{\Phi_C^{(k_n)}\left(\frac{\Omega_n}{\gamma_n}\right)}{\mathbb{E}\left(\Phi_C^{(k_n)}\left(\frac{\Omega_n}{\gamma_n}\right)\right)} \bullet \mathbb{P}_{\Omega_n}$$

Then, we still have

$$\frac{\mathbb{E}\left(x^{\Omega''_n}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)} = \frac{\mathbb{E}\left(\Phi_C^{(k_n)}\left(\frac{\Omega_n(x)}{\gamma_n}\right)\right)}{\mathbb{E}\left(\Phi_C^{(k_n)}\left(\frac{\Omega_n(1)}{\gamma_n}\right)\right)} \frac{\mathbb{E}\left(x^{\Omega_n}\right)}{\mathbb{E}\left(x^{P_{\gamma_n}}\right)}$$

and this last quantity converges locally uniformly to Φ_ω, i.e.

$$\left(\Omega''_n, H_n^{(\mathcal{P})}\right) \xrightarrow[n\to+\infty]{\text{mod-P}} \Phi_\omega$$

Now, we construct Ω''_n pathwise by means of a sequence of Bernoulli and uniform random variables.

Third step : Construction. Let $(B'_\ell)_\ell$ a sequence of independent $\{0,1\}$-Bernoulli random variables with $\mathbb{P}\left(B'_\ell = 1\right) = \frac{1}{\ell}$ independent of Ω_n. We have

$$\begin{aligned}
\mathbb{E}\left(x^{\Omega''_n}\right) &= \frac{1}{c_n}\mathbb{E}\left(x^{\Omega_n}\prod_{\ell=1}^{k_n}\left(1+\frac{1}{\ell}\left(\frac{\Omega_n}{\gamma_n}-1\right)\right)e^{-\frac{\Omega_n}{\gamma_n \ell}}\right)\\
&= \frac{1}{c'_n}\mathbb{E}\left((xv_n)^{\Omega_n}\prod_{\ell=1}^{k_n}\left(\frac{\Omega_n}{\gamma_n}\right)^{B'_\ell}\right) \quad \text{with } \ v_n := \exp\left(-H_{k_n}/\gamma_n\right)\\
&= \frac{1}{c''_n}\mathbb{E}\left((xv_n)^{\Omega_n}\prod_{\ell=1}^{k_n}\Omega_n^{B'_\ell(1/\gamma_n)}\right) \quad \text{with the notations of lemma 3.4.6}
\end{aligned}$$

Setting

$$C'_n := \sum_{\ell\leqslant k_n} B'_\ell(1/\gamma_n)$$

we get

$$\mathbb{E}\left(x^{\Omega''_n}\right) = \frac{1}{c_n}\mathbb{E}\left((xv_n)^{\Omega_n}\Omega_n^{C'_n}\right) = \frac{\mathbb{E}\left(v_n^{\Omega_n}\Omega_n^{C'_n}x^{\Omega_n}\right)}{\mathbb{E}\left(v_n^{\Omega_n}\Omega_n^{C'_n}\right)}$$

This random variable is the combination of two biases : a first exponential bias in the vein of lemma 3.4.6 with parameter v_n and a random iteration of size-bias transform, the number of times this transform is applied being given by C'_n. The effect of the exponential bias amounts to change the probabilities of Ω_n to get

$$\widetilde{\Omega}_n \stackrel{\mathcal{L}}{=} \sum_{k=1}^{\pi(n)} B_{\mathrm{p}_k}(v_n)$$

with $\pi(n) := \sum_{p\in\mathcal{P}} \mathbb{1}_{\{p\leqslant n\}} \sim \frac{n}{\log n}$ by the Prime Number Theorem, $\mathcal{P} := \{\mathrm{p}_k, k \geqslant 1\}$ and

$$\mathbb{E}\left(x^{\Omega''_n}\right) = \frac{\mathbb{E}\left(\widetilde{\Omega}_n^{C'_n} x^{\widetilde{\Omega}_n}\right)}{\mathbb{E}\left(\widetilde{\Omega}_n^{C'_n}\right)} = \frac{1}{\mathbb{E}\left(\widetilde{\Omega}_n^{C'_n}\right)} \sum_{\ell=0}^{k_n} \mathbb{P}\left(C'_n = \ell\right) \mathbb{E}\left(\widetilde{\Omega}_n^{\ell} x^{\widetilde{\Omega}_n}\right) \tag{3.24}$$

A size bias with a power ℓ is nothing else than the ℓ-th iteration of the usual size-bias transform defined in lemma 3.4.5, as one can see by writing, for a bounded measurable function f

$$\begin{aligned}
\frac{\mathbb{E}\left(X^2 f(X)\right)}{\mathbb{E}\left(X^2\right)} &=: \frac{1}{\mathbb{E}\left(X^2\right)}\mathbb{E}\left(Xg(X)\right) \quad \text{with } g(x) := xf(x) \\
&= \frac{\mathbb{E}\left(X\right)}{\mathbb{E}\left(X^2\right)}\mathbb{E}\left(g\left(X^{(s)}\right)\right) = \frac{\mathbb{E}\left(X\right)}{\mathbb{E}\left(X^2\right)}\mathbb{E}\left(X^{(s)} f\left(X^{(s)}\right)\right) \\
&= \frac{\mathbb{E}\left(X\right)}{\mathbb{E}\left(X^2\right)}\mathbb{E}\left(X^{(s)}\right)\mathbb{E}\left(f\left((X^{(s)})^{(s)}\right)\right)
\end{aligned}$$

and one can check setting $f = id : x \mapsto x$ in the definition of the size-bias transform that

$$\mathbb{E}\left(X^{(s)}\right) = \frac{\mathbb{E}\left(X^2\right)}{\mathbb{E}\left(X\right)}$$

i.e.

$$\mathbb{P}_{X^{(s,2)}} := \mathbb{P}_{(X^{(s)})^{(s)}} = \frac{X^2}{\mathbb{E}\left(X^2\right)} \bullet \mathbb{P}_X$$

From now on, we denote by $X^{(s,k)} := (X^{(s,k-1)})^{(s)}$ and $X^{(s,0)} := X$. In virtue of lemma 3.4.5, we have

$$\widetilde{\Omega}_n^{(s)} \stackrel{\mathcal{L}}{=} \widetilde{\Omega}_n - B_{\mathrm{p}_I}(v_n) + B_{\mathrm{p}_I}(v_n)^{(s)}$$

with $I \in [\![1, \pi(n)]\!]$ a random index independent of all random variables in presence of distribution

$$\mathbb{P}\left(I = k\right) = \frac{\frac{1}{\mathrm{p}_k + v_n + 1}}{\sum_{\ell=1}^{\pi(n)} \frac{1}{\mathrm{p}_\ell + v_n + 1}} \tag{3.25}$$

In addition, for a $\{0,1\}$-Bernoulli random variable B, we have

$$\mathbb{E}\left(x^{B^{(s)}}\right) = \frac{\mathbb{E}\left(Bx^B\right)}{\mathbb{E}\left(B\right)} = x$$

i.e. $B^{(s)} = 1$ almost surely (which amounts to change its parameter to 1). Hence,

$$\widetilde{\Omega}_n^{(s)} \stackrel{\mathcal{L}}{=} \widetilde{\Omega}_n - B_{\mathrm{p}_I}(v_n) + 1 = \sum_{k\leqslant \pi(n),\, k\neq I} B_{\mathrm{p}_k}(v_n) + 1$$

If we iterate the transformation, this amounts to toss a certain random variable J whose law is given by (3.25) independent of all random variables in presence and in particular independent of I. Two cases can occur : either $I = J$ in which case $\widetilde{\Omega}_n^{(s,2)} \stackrel{\mathcal{L}}{=} \widetilde{\Omega}_n^{(s)}$, or $I \neq J$ in which case $\widetilde{\Omega}_n^{(s,2)} \stackrel{\mathcal{L}}{=} \sum_{k\neq I,J} B_{\mathrm{p}_k}(v_n) + 2$, which we can sumarize into

$$\widetilde{\Omega}_n^{(s,2)} \stackrel{\mathcal{L}}{=} \sum_{k\neq I,J} B_{\mathrm{p}_k}(v_n) + 1 + \mathbb{1}_{\{I\neq J\}}$$

The third iterate gives

$$\widetilde{\Omega}_n^{(s,3)} \stackrel{\mathcal{L}}{=} \sum_{k\neq I_1,I_2,I_3} B_{\mathrm{p}_k}(v_n) + \delta(I_1, I_2, I_3)$$

with

$$\delta(I_1, I_2, I_3) = \begin{cases} 1 & \text{if } I_1 = I_2 = I_3 \\ 2 & \text{if } I_i = I_j \neq I_k \quad \text{for } \{i,j,k\} = \{1,2,3\} \\ 3 & \text{if } I_1 \neq I_2 \neq I_3 \neq I_1 \end{cases}$$

At the ℓ-th iteration, one has, with a sequence of i.i.d. indexes $(I_\ell)_\ell$ of law given by (3.25)

$$\widetilde{\Omega}_n^{(s,\ell)} \stackrel{\mathcal{L}}{=} \sum_{k\neq I_1,\dots,I_\ell} B_{\mathrm{p}_k}(v_n) + \delta(I_1, \dots, I_\ell)$$

where $\delta(I_1, \dots, I_\ell)$ is the lenght of the random partition $\lambda \vdash \ell$ constructed by means of the following "paintbox", i.e. define the equivalence relation by

$$k \sim r \quad \Longleftrightarrow \quad I_k = I_r$$

Then, the equivalence classes of this relation define a random partition. In the case where $I \sim \mathscr{U}\left(\llbracket 1, \pi(n) \rrbracket\right)$, this random partition is equal in law to the cycle structure of a random uniform permutation $\sigma \in \mathfrak{S}_\ell$ and in particular, $\delta(I_1, \dots, I_\ell) = C(\sigma)$. In the case of our indexes, this distribution has still to be precised.

Last, the equality (3.24) is equivalent to

$$\Omega_n'' \stackrel{\mathcal{L}}{=} \widetilde{\Omega}_n^{(s,C_n')}$$

which implies that

$$\Omega_n'' \stackrel{\mathcal{L}}{=} \sum_{k\neq I_1,\dots,I_{C_n'}} B_{\mathrm{p}_k}(v_n) + \delta(I_1, \dots, I_{C_n'}) \quad \text{with} \quad C_n' := \sum_{\ell\leqslant k_n} B_\ell'(1/\gamma_n)$$

all the random variables considered being independent.

$\square$

Remark 3.5.2. One can compare Ω''_n with (3.23) : in addition to have one degree of freedom left in the former model, such a construction gives an independant model for $\omega(U_n)$; in the latest construction, the random variables involved are not independant anymore.

Remark 3.5.3. Note the following interpretation of the corrective term : one has refined the Erdös-Kac model Ω_n by imposing a certain proportion of primes (in quantity $\delta(I_1, \dots, I_{C'_n})$) to be divisors with probability one. If one knows that a certain amount of primes are divisors, the least is to avoid them in the set of primes to consider for counting the prime divisors. This operation can hence be (almost) understood as a conditioning (up to the change of probabilities) :

$$\widehat{\Omega}_n \stackrel{\mathcal{L}}{=} \left(\Omega_n(v_n)\middle| B_{\mathrm{p}_{I_1}}(v_n) = 1, \dots, B_{\mathrm{p}_{I_{C'_n}}}(v_n) = 1\right) \stackrel{\mathcal{L}}{\approx} \left(\Omega_n \middle| B_{\mathrm{p}_{I_1}} = 1, \dots, B_{\mathrm{p}_{I_{C'_n}}} = 1\right)$$

Nevertheless, a drawback of this model is the fact that the primes randomly selected are not the large ones ; they are selected at random in the whole interval $[\![1, \pi(n)]\!]$ and not above a certain treshold, and due to the structure of the law (3.25), these are the small primes that are cancelled in the sum.

Remark 3.5.4. An abstract form of this result can be given in the general framework of a product of two $\{0,1\}$-Bernoulli mod-Poisson limiting functions given in (3.1), and applies in particular to the case of $\omega_q(P_n)$ of the example 3.3.4.

3.5.3 Perspectives and open problems

The latest review of mod-* convergence has settled several questions and opened several directions of investigation.

A first natural question concerns the total dependency case given by the Salem-Zygmund theorem or the Fortet theorem given in [40], i.e. with one single random variables and functions having a certain orthogonality property ; in particular, the case of functions with values in $\{0,1\}$ would lead to an interesting mod-Poisson case. The limiting function occuring in such a case could be interesting to analyse and in particular to compare with the case of $\omega(U_n)$, since all the limiting functions in mod-Poisson models lead to a product phenomena involving the general type (3.1).

Concerning $\omega(U_n)$, it would be interesting to go beyond the Erdös-Turan theorem and to look at a functional renormalisation, i.e. the Erdös-Kubilius theorem. Such a functional renormalisation gives at the limit a Brownian motion, but a more refined one could give a Poisson process. Note that the functional generalisation of mod-* convergence in the functional setting by means of a functional Fourier or Laplace transform is straightforward.

In the case of the moments conjecture, the construction of a random variable micmicking the mod-Gaussian fluctuations (i.e. converging in the mod-Gaussian sense to the limiting function of the Zeta) still remains to be done.

More generally, a better construction of the mod-Poisson model for $\omega(U_n)$ has to be done. A general guess would be a random variable of the type

$$\sum_{k \neq I_1, \dots, I_{Z_n}} B_{\mathrm{p}_k} + \delta(I_1, \dots, I_{Z_n})$$

with the $(I_\ell)_\ell$ independent uniform random variables on $[\![A, \pi(n)]\!]$ where A is to be found, and Z_n a random integer to be found. The advantages of such a model is the absence of independence between the two random variables in presence, the natural apparition of the number of cycles with the paintbox process (since, with the I_ℓ's uniform, this is the number of cycles under the Haar measure of a certain symmetric group), and the interpretation in terms of conditioning on the large primes.

3.6 Mod-Poisson convergence with auxiliary randomization

3.6.1 An interesting fact

One striking phenomenon occuring with the limiting functions (3.1) that appear in the mod-Poisson convergence of Bernoulli models is the fact that they are inverses of Mellin transforms. This is clear in the case of the corrective Gamma factor in $\omega(U_n)$ as remarked after equation (3.3), but in the general case, one has

$$\frac{1}{\Phi(x)} = \prod_{k\geqslant 1}(1 + p_k(x-1))^{-1}e^{p_k(x-1)} = \prod_{k\geqslant 1}\mathbb{E}\left(e^{-p_k(x-1)\mathbf{e}^{(k)}}\right)e^{p_k(x-1)} = \mathbb{E}\left(Z^{x-1}\right)$$

with, if $(\mathbb{e}^{(k)})_k$ denotes a sequence of i.i.d. exponentially distributed random variables,

$$Z \stackrel{\mathcal{L}}{=} \exp\left(\sum_{k\geqslant 1} p_k(1-\mathbb{e}^{(k)})\right)$$

This last sum converges in L^2 since $\sum_k p_k^2 < \infty$. The case where $p_k = 1/k$ corresponds to $Z \stackrel{\mathcal{L}}{=} \mathbb{e}$, which amounts to the identity (3.4).

Now, consider a sequence of random variables $(X_n)_n$ with values in $\mathbb{N}$ and converging in the mod-Poisson sense at speed $(\gamma_n)_n$ to a function Φ_X that splits into $\Phi_X = \Phi_M\Phi_B$ where Φ_M is the limiting function corresponding to a sequence $(M_n)_n$ (a "model" of $(X_n)_n$ that one can think of as an "independent model", although this last vision is restrictive and unnecessary here) and Φ_B is a correction corresponding e.g. to a Bernoulli model, or more generally being equal to the inverse of a Mellin function.

Supposing that $\gamma_n = \log n$, and taking $P(\log n) \sim \mathscr{P}(\log n)$, one has locally uniformly

$$\frac{\mathbb{E}\left(x^{X_n}\right)}{\mathbb{E}\left(x^{P(\log n)}\right)} \xrightarrow[n\to+\infty]{} \Phi_X(x) = \Phi_M(x)\Phi_B(x)$$

Since there exists a random variable Z such that $\Phi_B(x) = 1/\mathbb{E}\left(Z^{x-1}\right)$, one can write, with Z independent of $P(\log n)$,

$$\mathbb{E}\left(x^{P(\log(nZ))}\right) = \mathbb{E}\left(e^{(x-1)\log(nZ)}\right) = e^{(x-1)\log(n)}\mathbb{E}\left(Z^{(x-1)}\right) = \frac{\mathbb{E}\left(x^{P(\log(n))}\right)}{\Phi_B(x)}$$

which implies

$$\frac{\mathbb{E}\left(x^{X_{nZ}}\right)}{\mathbb{E}\left(x^{P(\log(n))}\right)} = \frac{\mathbb{E}\left(x^{X_{nZ}}\right)}{\mathbb{E}\left(x^{P(\log(nZ))}\right)}\frac{\mathbb{E}\left(x^{P(\log(nZ))}\right)}{\mathbb{E}\left(x^{P(\log(n))}\right)} \xrightarrow[n\to+\infty]{} \Phi_M(x)$$

In fact, every Mellin transform of a random variable $Y \geqslant 1$ can be understood as the characteristic function of a randomised Poisson random variable with the identity

$$\mathbb{E}\left(Y^{x-1}\right) = \mathbb{E}\left(e^{(x-1)\log Y}\right) = \mathbb{E}\left(x^{P(\log Y)}\right)$$

Thus, by an additional randomization, one finds the independent model. If one supposes that $\gamma_n = \log\log n$ as in the case of $\omega(U_n)$, the same reasoning can be done using $P(\log\log(n^Z))$ and X_{n^Z}. In the general case, one can define the randomization by setting

$$\gamma_n(T) \equiv \gamma_n + T$$

which amounts to write $\gamma_n(T) := \gamma_{N(n,T)}$ with

$$N(n,T) = \gamma^{-1}_{\gamma_n+T} := \inf\left\{k \geqslant 0 \ / \ \gamma_k > \gamma_n + T\right\}$$

Note that one can adapt the last reasonning by setting an additional error term in the random variable Z, i.e. use Z_n that converges in distribution to Z at a certain rate.

All these considerations tackle the question of finding an additional randomization in sequences of dependent random variables converging in the mod-* sense with an inverse Mellin transform as a corrective function that could explain the splitting phenomenon, in particular in the case of $\omega(U_n)$ since

$$\frac{\mathbb{E}\left(x^{\omega(U_{n^{\mathbf{e}}})}\right)}{\mathbb{E}\left(x^{P(\log\log(n))}\right)} \xrightarrow[n\to+\infty]{} \Phi_\Omega(x) \tag{3.26}$$

3.6.2 Randomization on the Symmetric group

Several examples in Probability theory illustrate the fact that a certain additional randomization of a particular parameter in a sequence of random variables reveals one of its fundamental aspect, hidden at first glance. This is for instance the case of a random partition selected according to the Plancherel distribution $\mathcal{P}_n(\lambda) := d_\lambda^2/n!$: setting $n \sim \mathscr{P}(t)$ for $t > 0$ gives the Poisson-Plancherel distribution which is a measure on the infinite Young graph, or, equivalently, on an infinite number of particules. The point process is then determinantal (see e.g. [80]) and one can recover the asymptotic behaviour of the initial Plancherel point process by a derandomization, a de-Poissonisation in this case (see e.g. [80, 56]).

A simpler example can be given with the cycle structure of a random uniform permutation : Polya's cycle index theorem (see e.g. [71], p. 25) expresses in terms of Laplace transform the following identity in law

$$\mathcal{L}_{\mathbb{P}_\theta^{(n)}}(c_1, \dots, c_n) = \mathcal{L}\left(P_1, \dots, P_n \middle| \sum_{k=1}^n kP_k = n\right) \tag{3.27}$$

with $(P_1, \dots, P_n)$ a vector of independent coordinates of law

$$P_k \sim \mathscr{P}\left(\frac{\theta}{k}\right)$$

Indeed, using the multi-index notation $x^\alpha := \prod_{k\geqslant 1} x_k^{\alpha_k}$, we define the *cycle indicator* or *cycle index polynomial* of a subgroup G of $\mathfrak{S}_n$ as the symmetric function given by

$$\mathrm{CIP}(G)(x) := \frac{1}{|G|}\sum_{g\in G} x^{\boldsymbol{C}(g)}$$

Noting $\mathcal{U}(G)$ the uniform measure on G, we see that

$$\mathrm{CIP}(G) = \mathbb{E}_{\mathcal{U}(G)}\big(x^{\boldsymbol{C}}\big) \tag{3.28}$$

Hence, the cycle index polynomial of G is the multivariate probability generating function, that is, the generating function of the random variable $\boldsymbol{C}$.

In the case where $G = \mathfrak{S}_n$, we have (see e.g. [71], example 9 p. 29) using the convention $x_\lambda := \prod_{k\geqslant 1} x_{\lambda_k}$ for $\lambda \vdash n$,

$$\mathrm{CIP}(\mathfrak{S}_n) := \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} x^{\boldsymbol{C}(\sigma)} = \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} x_{\mathrm{ct}(\sigma)} = \sum_{\lambda\vdash n}\frac{x_\lambda}{z_\lambda} = h_n$$

Here x_λ plays the rôle of the extitpower functions usually denoted by p_λ in [71]. The fact that

$$H(t) := \sum_{n\geqslant 1} h_n t^n = \exp\left(\sum_{k\geqslant 1}\frac{x_k t^k}{k}\right)$$

is Polya's CIP theorem.

This last equality has the following probabilistic interpretation : multiplying both sides by $(1-t)$, one has

$$\sum_{n\geqslant 1}(1-t)t^n\mathbb{E}_{\mathcal{U}(\mathfrak{S}_n)}\big(x^{\boldsymbol{C}}\big) = (1-t)\exp\left(\sum_{k\geqslant 1}\frac{x_k t^k}{k}\right) = \exp\left(\sum_{k\geqslant 1}\frac{(x_k-1)t^k}{k}\right) = \prod_{k\geqslant 1} e^{(x_k-1)\frac{t^k}{k}}$$

Let G_t be a Geometrically distributed random variable, i.e. $\mathbb{P}(G_t = k) = t^k(1-t)$ for all $k\in\mathbb{N}$ and let $(P(t^k/k))_k$ be a sequence of independent random variables Poisson-distributed with parameter t^k/k. Then,

$$\mathbb{E}_{\mathcal{U}(\mathfrak{S}_{G_t})}\big(x^{\boldsymbol{C}}\big) := \sum_{n\geqslant 1}\mathbb{P}(G_t = n)\,\mathbb{E}_{\mathcal{U}(\mathfrak{S}_n)}\big(x^{\boldsymbol{C}}\big) = \prod_{k\geqslant 1}\mathbb{E}\left(x_k^{P(t^k/k)}\right) \tag{3.29}$$

Noting $c_k(\sigma)$ the number of k-cycles of $\sigma\in\mathfrak{S}_n$, we see that under the geometrisation $\mathcal{U}(\mathfrak{S}_{G_t})$ of the uniform measure, the cycle structure $(c_k)_{k\geqslant 1}$ has the same distribution as the sequence of independent random variables $(P(t^k/k))_k$

$$(c_k)_{k\geqslant 1} \stackrel{\mathcal{L}}{=} \left(P(t^k/k)\right)_{k\geqslant 1}$$

In the same maneer as the Poisson-Plancherel measure highlights the determinantal structure of a random partition selected according to it, this last Geometric-Uniform measure highlights the character independent Poisson of the cycle structure of a random permutation selected according to it.

One can remark that G_t can be expressed by means of the sequence $\big(P(t^k/k)\big)_k$ via

$$\big(G_t, (c_k)_{k\geqslant 1}\big) \stackrel{\mathcal{L}}{=} \left(\sum_{k\geqslant 1} kP(t^k/k), \Big(P(t^k/k)\Big)_{k\geqslant 1}\right)$$

Letting $t \to 1$ in this last identity gives back (3.27).

There are several analogies between permutations and primes. Some are listed in [2], pp. 32 or in [47]. In particular, permutations and integers share a divisibility property and a prime decomposition :

$$\forall\, \sigma \in \mathfrak{S}_n, \quad \sigma = \prod_{c\in\mathcal{C}(\sigma)} c$$
$$\forall\, n \in \mathbb{N}, \quad n = \prod_{p\in\mathcal{P}} p^{v_p(n)}$$

Here, $\mathcal{C}(\sigma)$ is the set of cycles of σ, the convention on the product being defined in section 1.1.1, and $\boldsymbol{v} = (v_p)_{p\in\mathcal{P}}$ is the *valuation structure*, analogous to the cycle structure $\boldsymbol{C} = (c_k)_k$. The question of an identity analogous to Polya's CIP theorem is thus relevant.

Polya's CIP theorem deals with

$$\sum_{n\geqslant 1} \mathbb{P}\left(G_t = n\right) \mathbb{E}_{\mathcal{U}(\mathfrak{S}_n)}\big(x^{\boldsymbol{C}}\big)$$

and the analogous quantity here would be, with a certain random variable G'_t to be discovered and with $x^{\boldsymbol{v}} := \prod_{p\in\mathcal{P}} x_p^{v_p}$

$$\sum_{n\geqslant 1} \mathbb{P}\left(G'_t = n\right) \mathbb{E}\left(x^{\boldsymbol{v}(U_n)}\right)$$

3.6.3 The delta-Zeta distribution

Definition

Definition 3.6.1 (Delta-Zeta distribution)**.** Let $\alpha > 1$. The *delta-Zeta distribution* denoted by $\boldsymbol{\delta\zeta}(\alpha)$ is the distribution on $\mathbb{N}^*$ defined by, for $\delta_\alpha \sim \boldsymbol{\delta\zeta}(\alpha)$

$$\forall\, k \geqslant 1, \qquad \mathbb{P}\left(\delta_\alpha = k\right) := \frac{k}{\zeta(\alpha)}\left(\frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha}\right)$$

This is a probability distribution : it is positive since $x \mapsto 1/x^\alpha$ is decreasing and

$$\sum_{k\geqslant 1} \frac{k}{\zeta(\alpha)}\left(\frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha}\right) = 1$$

as one can see by performing an Abel summation (integration by parts), that is (setting $u_0 := 0$)

$$\sum_{k\geqslant 1} u_k(v_k - v_{k+1}) = \sum_{k\geqslant 1} u_k v_k - \sum_{k\geqslant 1} u_k v_{k+1} = \sum_{k\geqslant 1} u_k v_k - \sum_{k\geqslant 1} u_{k-1} v_k = \sum_{k\geqslant 1} (u_k - u_{k-1}) v_k$$

The application of this formula with $u_k = k$ and $v_k = k^{-\alpha}$ gives the result.

The Laplace transform of this sequence is given by, for all $x \in]0,1[$,

$$\begin{aligned}\mathbb{E}\left(x^{\delta_\alpha}\right) &:= \frac{1}{\zeta(\alpha)}\sum_{k\geqslant 1} kx^k\left(\frac{1}{k^\alpha}-\frac{1}{(k+1)^\alpha}\right)\\ &= \frac{1}{\zeta(\alpha)}\sum_{k\geqslant 1}\left(kx^k-(k-1)x^{k-1}\right)\frac{1}{k^\alpha}\end{aligned}$$

Define the polylogarithm function by

$$\mathrm{Li}_\alpha(z) := \sum_{k\geqslant 1}\frac{z^k}{k^\alpha}$$

Then,

$$\mathbb{E}\left(x^{\delta_\alpha}\right) = \frac{1}{\zeta(\alpha)}\left(\left(1-\frac{1}{x}\right)\mathrm{Li}_{\alpha-1}(x)+\frac{1}{x}\,\mathrm{Li}_\alpha(x)\right)$$

Last, remark that δ_α can be defined as a size-bias distribution by setting

$$\mathbb{P}\left(\delta_\alpha = k\right) = \frac{k}{\zeta(\alpha)}\mathbb{P}\left(R_\alpha = k\right) := \frac{k}{\zeta(\alpha)}\left(\frac{1}{k^\alpha}-\frac{1}{(k+1)^\alpha}\right)$$

with, $\mathbb{e}$ denoting a random variable exponentially distributed and $[x]$ the integer part of x,

$$R_\alpha \stackrel{\mathcal{L}}{=} \left[\exp\left(\mathbb{e}/\alpha\right)\right]$$

One can easily check that $\mathbb{P}\left(\left[\exp\left(\mathbb{e}/\alpha\right)\right] = k\right) = k^{-\alpha}-(k+1)^{-\alpha}$ and $\mathbb{E}\left(R_\alpha\right) = \zeta(\alpha)$.

Fluctuations when $\alpha \to 1$

One has

$$\mathbb{P}\left(\delta_\alpha \leqslant x\right) = \frac{1}{\zeta(\alpha)}\sum_{k=1}^{x} k\left(\frac{1}{k^\alpha}-\frac{1}{(k+1)^\alpha}\right) = \frac{1}{\zeta(\alpha)}\sum_{k=1}^{x}\frac{1}{k^\alpha}-\frac{1}{\zeta(\alpha)(x+1)^\alpha}$$

by an Abel summation. In particular, for all $t \geqslant 0$, one has

$$\begin{aligned}\mathbb{P}\left(\delta_\alpha \leqslant e^{\frac{t}{\alpha-1}}\right) &= \frac{1}{\zeta(\alpha)}\sum_{k=1}^{e^{\frac{t}{\alpha-1}}}\frac{1}{k^\alpha}-\frac{1}{\zeta(\alpha)\left(e^{\frac{t}{\alpha-1}}+1\right)^\alpha}\\ &= \frac{1}{\zeta(\alpha)}\int_1^{e^{\frac{t}{\alpha-1}}}\frac{dx}{x^\alpha}+O\left(\frac{1}{\zeta(\alpha)\ e^{\frac{t(-\alpha+1)}{\alpha-1}}}\right)+O\left(\frac{e^{-\frac{\alpha t}{\alpha-1}}}{\zeta(\alpha)}\right)\\ &= \frac{1}{\zeta(\alpha)}\frac{e^{-t}-1}{-\alpha+1}+O\left(\frac{e^t}{\zeta(\alpha)}\right)\ \xrightarrow[\alpha\to 1]{} 1-e^{-t} = \mathbb{P}\left(\mathbb{e}\leqslant t\right)\end{aligned}$$

where $\mathbb{e}$ denotes a random variable exponentially distributed and where we have used the fact that

$$(\alpha-1)\zeta(\alpha) \xrightarrow[\alpha\to 1]{} 1 \tag{3.30}$$

Thus, we have the convergence in law

$$(\alpha-1)\log\delta_\alpha \xrightarrow[\alpha\to 1]{\mathcal{L}} \mathbb{e}$$

The Zeta distribution

A random variable $Z_\alpha : \Omega \to \mathbb{N}^*$ is said to be $\boldsymbol{\zeta}(\alpha)$-distributed for a certain $\alpha > 1$ if its distribution is given by

$$\mathbb{P}\left(Z_\alpha = k\right) := \frac{1}{\zeta(\alpha)k^\alpha}$$

One has

$$\begin{aligned}\mathbb{E}\left(Z_\alpha^{is}\right) &= \frac{1}{\zeta(\alpha)}\sum_{k\geqslant 1}\frac{k^{is}}{k^\alpha} = \frac{\zeta(\alpha-is)}{\zeta(\alpha)} = \prod_{p\in\mathcal{P}}\frac{1-p^{-\alpha}}{1-p^{-\alpha+is}}\\ &= \prod_{p\in\mathcal{P}}\mathbb{E}\left(p^{is\,\mathcal{G}(p^{-\alpha})}\right)\end{aligned}$$

where $\mathcal{G}(q)$ is a random variable with a Geometric distribution given by

$$\mathbb{E}\left(x^{\mathcal{G}(q)}\right) = \frac{1-q}{1-qx}$$

We hence have

$$Z_\alpha \stackrel{\mathcal{L}}{=} \exp\left(\sum_{p\in\mathcal{P}}\log p\ \mathcal{G}\left(p^{-\alpha}\right)\right)$$

the $\mathcal{G}\left(p^{-\alpha}\right)$ being independent. Note that since the geometric random variables are infinitely divisible, $\log Z_\alpha$ is also infinitely divisible, a result that dates back to Kintchine in 1938.

This distribution is a particular case of *multiplicative distribution*, i.e. a distribution for which the valuation structure of the random variable is composed with independant random variables : $\boldsymbol{v}(Z_\alpha)$ is a random vector for $Z_\alpha \sim \boldsymbol{\zeta}(\alpha)$ and one has, for $\boldsymbol{x} = (x_p)_{p\in\mathcal{P}}$

$$\mathbb{E}\left(\boldsymbol{x}^{\boldsymbol{v}(Z_\alpha)}\right) = \frac{1}{\zeta(\alpha)}\sum_{k\geqslant 1}\frac{\boldsymbol{x}^{\boldsymbol{v}(k)}}{k^\alpha} = \prod_{p\in\mathcal{P}}\frac{1-p^{-\alpha}}{1-p^{-\alpha}\boldsymbol{x}^{\boldsymbol{v}(p)}} = \prod_{p\in\mathcal{P}}\frac{1-p^{-\alpha}}{1-p^{-\alpha}x_p}$$

since $f : k \mapsto \boldsymbol{x}^{\boldsymbol{v}(k)}$ is a multiplicative function (that is $f(k\,\ell) = f(k)f(\ell)$ for all $k, \ell \in \mathbb{N}^*$) due to the fact that

$$v_p(k\,\ell) = v_p(k) + v_p(\ell)$$

We have in addition used the fact that for all $p, q \in \mathcal{P}$,

$$v_p(q) = \mathbb{1}_{\{p=q\}}$$

and the following formula valid for a multiplicative function $f : \mathbb{N}^* \to \{z \in \mathbb{C} \ |z| < 1\}$ such that $\sum_k |f(k)| < \infty$

$$\sum_{k \geqslant 1} f(k) = \prod_{p \in \mathcal{P}} \frac{1}{1 - f(p)} \tag{3.31}$$

A simple proof of this last formula consists in writing, with $\mathcal{P} = \{p_\ell, \ \ell \geqslant 1\}$

$$\begin{aligned}
\prod_{p \in \mathcal{P}} \frac{1}{1 - f(p)} &= \prod_{p \in \mathcal{P}} \sum_{k_p \geqslant 0} f(p)^{k_p} \\
&= \sum_{n \geqslant 1} \sum_{k_{p_1}, k_{p_2}, \ldots, k_{p_n} \geqslant 0} \prod_{\ell=1}^{n} f(p_\ell)^{k_{p_\ell}} \\
&= \sum_{n \geqslant 1} \sum_{k_{p_1}, k_{p_2}, \ldots, k_{p_n} \geqslant 0} f\left(\prod_{\ell=1}^{n} p_\ell^{k_{p_\ell}}\right) \quad \text{by multiplicativity} \\
&= \sum_{n \geqslant 1} f(n)
\end{aligned}$$

We thus see that

$$\mathbb{E}\left(\boldsymbol{x}^{\boldsymbol{v}(Z_\alpha)}\right) = \prod_{p \in \mathcal{P}} \mathbb{E}\left(x_p^{\mathcal{G}(p^{-\alpha})}\right)$$

and more generally, if f is a real positive multiplicative function satisfying the last hypotheses, defining the *multiplicative distribution* $\mathfrak{M}(f)$ by, for $X_f \sim \mathfrak{M}(f)$

$$\mathbb{P}(X_f = k) = \frac{f(k)}{\sum_{\ell \geqslant 1} f(\ell)}$$

one has for the same reasons as before

$$\mathbb{E}\left(\boldsymbol{x}^{\boldsymbol{v}(X_f)}\right) = \prod_{p \in \mathcal{P}} \mathbb{E}\left(x_p^{\mathcal{G}(f(p))}\right)$$

and in particular

$$X_f \stackrel{\mathcal{L}}{=} \exp\left(\sum_{p \in \mathcal{P}} \log p \ \mathcal{G}(f(p))\right)$$

Remark 3.6.2. One has

$$\mathbb{P}(Z_\alpha \leqslant n) = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{n} \frac{1}{k^\alpha} = \mathbb{P}(\delta_\alpha \leqslant n) + \frac{1}{\zeta(\alpha)(n+1)^\alpha}$$

hence

$$(\alpha - 1) \log Z_\alpha \xrightarrow[\alpha \to 1]{\mathcal{L}} \mathbb{e}$$

The analogue of Polya's CIP

The utility of the delta-Zeta distribution comes from the following theorem

Theorem 3.6.3. *Let $U_n \sim \mathscr{U}(\llbracket 1, n \rrbracket)$ be a uniform random variable in $\llbracket 1, n \rrbracket$ and $\delta_\alpha \sim \boldsymbol{\delta\zeta}(\alpha)$. Suppose that δ_α is independent of U_n. Then*

$$U_{\delta_\alpha} \sim \boldsymbol{\zeta}(\alpha)$$

Moreover, if there exists a couple (X, Y) of integer-valued random variables such that X is independent of U_n and $Y \sim \mathfrak{M}(f)$ for a certain multiplicative function f, then, there exists a unique $\alpha > 1$ such that $f(k) = k^{-\alpha}$ for all $k \geqslant 1$.

Proof. For $s \in \mathbb{R}$, one has

$$\begin{aligned}
\mathbb{E}\left(U_{\delta_\alpha}^{is}\right) &= \sum_{k \geqslant 1} \mathbb{P}\left(\delta_\alpha = k\right) \mathbb{E}\left(U_k^{is}\right) \\
&= \frac{1}{\zeta(\alpha)} \sum_{k \geqslant 1} k \left(\frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha} \right) \frac{1}{k} \sum_{\ell=1}^{k} \ell^{is} \\
&= \frac{1}{\zeta(\alpha)} \sum_{k \geqslant 1} \frac{k^{is}}{k^\alpha} \quad \text{by Abel summation} \\
&= \mathbb{E}\left(Z_\alpha^{is}\right) \quad \text{with} \quad Z_\alpha \sim \boldsymbol{\zeta}(\alpha)
\end{aligned}$$

One concludes by injectivity of the Fourier-Mellin transform.

Now, consider a couple (X, Y) defined by the subordination equation $U_X \stackrel{\mathcal{L}}{=} Y \sim \mathfrak{M}(f)$. Then,

$$\begin{aligned}
\mathbb{E}\left(U_X^{is}\right) &= \sum_{k \geqslant 1} \mathbb{P}\left(X = k\right) \frac{1}{k} \sum_{\ell=1}^{k} \ell^{is} = \sum_{k \geqslant 1} k^{is} \left(\sum_{\ell \geqslant k} \frac{1}{\ell} \mathbb{P}\left(X = \ell\right) \right) \\
\mathbb{E}\left(Y^{is}\right) &= \sum_{k \geqslant 1} k^{is} \frac{f(k)}{\sum_{\ell \geqslant 1} f(\ell)}
\end{aligned}$$

This implies that for all $k \geqslant 1$,

$$\sum_{\ell \geqslant k} \frac{1}{\ell} \mathbb{P}\left(X = \ell\right) = \frac{f(k)}{\sum_{\ell \geqslant 1} f(\ell)}$$

Solving this linear equation in $\mathbb{P}\left(X = k\right)$ gives

$$\mathbb{P}\left(X = k\right) = \frac{k}{\sum_{\ell \geqslant 1} f(\ell)} \left(f(k) - f(k+1)\right)$$

It is clear that

$$\sum_{k \geqslant 1} \frac{k}{\sum_{\ell \geqslant 1} f(\ell)} \left(f(k) - f(k+1)\right) = 1$$

so, in order to have a probability distribution, one only needs to have $f(k) - f(k+1) \geqslant 0$. But a classical theorem on extremal order of arithmetic functions (see e.g. [98], pp. 82) characterizes the only multiplicative decreasing functions on $\mathbb{N}^*$ as being the power functions, i.e. there exists a unique $\alpha > 1$ such that

$$f(k) = k^{-\alpha}$$

□

As a direct corollary of this last theorem, one has the following "Arithmetic CIP" formula

$$\sum_{n \geqslant 1} \mathbb{P}(\delta_\alpha = n)\, \mathbb{E}\left(x^{\boldsymbol{v}(U_n)}\right) = \prod_{p \in \mathcal{P}} \mathbb{E}\left(x_p^{\mathcal{G}(p^{-\alpha})}\right) \tag{3.32}$$

analogous to Polya's CIP theorem (3.29) (using $\sigma_n \sim \mathrm{Haar}(\mathfrak{S}_n)$ to strengthen the analogy)

$$\sum_{n \geqslant 1} \mathbb{P}(\mathcal{G}(t) = n)\, \mathbb{E}\left(x^{\boldsymbol{C}(\sigma_n)}\right) = \prod_{k \geqslant 1} \mathbb{E}\left(x_k^{P(t^k/k)}\right)$$

Discussion

One has

$$\omega(Z_\alpha) = \sum_{p \in \mathcal{P}} \mathbb{1}_{\{v_p(Z_\alpha) \geqslant 1\}} = \sum_{p \in \mathcal{P}} \mathcal{B}\left(p^{-\alpha}\right)$$

with $(\mathcal{B}(p^{-\alpha}))_p$ a sequence of independent Bernoulli random variables of parameter $p^{-\alpha}$. In particular, one has

$$\mathbb{E}\left(x^{\omega(Z_\alpha)}\right) = \prod_{p \in \mathcal{P}} \left(1 + \frac{x-1}{p^\alpha}\right) = \exp\left((x-1)\sum_{p \in \mathcal{P}} \frac{1}{p^\alpha}\right) \prod_{p \in \mathcal{P}} \left(1 + \frac{x-1}{p^\alpha}\right) e^{-\frac{x-1}{p^\alpha}}$$

Define for $\alpha > 1$

$$\zeta_{\mathcal{P}}(\alpha) := \sum_{p \in \mathcal{P}} \frac{1}{p^\alpha}$$

We have

$$\log \zeta(\alpha) = \sum_{k \geqslant 1} \frac{\zeta_{\mathcal{P}}(k\alpha)}{k} = \zeta_{\mathcal{P}}(\alpha) + \sum_{k \geqslant 2} \frac{\zeta_{\mathcal{P}}(k\alpha)}{k}$$

Using (3.30), one deduces that

$$\zeta_{\mathcal{P}}(\alpha) \underset{\alpha \to 1^+}{\sim} -\log(\alpha - 1) = \log\left(\frac{1}{\alpha - 1}\right)$$

In particular, locally uniformly in $x > 0$,

$$\frac{\mathbb{E}\left(x^{\omega(Z_\alpha)}\right)}{\mathbb{E}\left(x^{P(\log(1/(\alpha-1)))}\right)} \xrightarrow[\alpha \to 1]{} \prod_{p \in \mathcal{P}} \left(1 + \frac{x-1}{p}\right) e^{-\frac{x-1}{p}} = \Phi_\Omega(x)$$

In order to choose a α_n that mimics the mod-Poisson convergence of $\omega(U_n)$, one thus needs to take $1/(\alpha_n - 1) = \log n$, which gives

$$\alpha_n := 1 + \frac{1}{\log n}$$

Since

$$\delta_\alpha \underset{\alpha \to 1}{\overset{\mathcal{L}}{\approx}} \exp\left(\frac{\mathbb{e}}{\alpha - 1}\right)$$

with this choice of α_n, one has

$$\delta_{\alpha_n} \underset{n \to +\infty}{\overset{\mathcal{L}}{\approx}} n^{\mathbb{e}}$$

$$Z_{\alpha_n} \overset{\mathcal{L}}{=} U_{\delta_{\alpha_n}} \underset{n \to +\infty}{\overset{\mathcal{L}}{\approx}} U_{n^{\mathbb{e}}}$$

This explains the convergence (3.26) and the apparition of the Γ factor as the correction to the independent model.

Note that the reasoning performed here is general to all models sharing such a property of giving an independent structure by means of a randomisation, like the cycle structure of a random uniform (or Ewens) permutation, and to all "separable" statistics, i.e. in the arithmetic case, the statistics that write as a sum $\sum_p f_p(v_p)$ such as ω (take $f_p(x) = \mathbb{1}_{\{x \geqslant 1\}}$).

Example 3.6.4. The total number of prime divisors of an integer $N \in \mathbb{N}^*$ is defined by

$$\Omega(N) := \sum_{p \in \mathcal{P}} v_p(N)$$

Selberg proved in [88] the mod-Poisson convergence

$$\frac{\mathbb{E}\left(x^{\Omega(U_n)}\right)}{\mathbb{E}\left(x^{P(\log\log n)}\right)} \xrightarrow[n \to +\infty]{} \frac{1}{\Gamma(x)} \prod_{p \in \mathcal{P}} \left(1 - \frac{x}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^{x}$$

Writing

$$\begin{aligned}
\prod_{p \in \mathcal{P}} \left(1 - \frac{x}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^{x} &= \prod_{p \in \mathcal{P}} \frac{1 - \frac{1}{p}}{1 - \frac{x}{p}} e^{(x-1)\log\left(1 - \frac{1}{p}\right)} \\
&= e^{(-\gamma + \kappa_{\mathcal{P}})(x-1)} \prod_{p \in \mathcal{P}} \frac{1 - \frac{1}{p}}{1 - \frac{x}{p}} e^{-\frac{x-1}{p}}
\end{aligned}$$

one sees that

$$\frac{\mathbb{E}\left(x^{\Omega(U_n)}\right)}{\mathbb{E}\left(x^{P(H_n^{(\mathcal{P})})}\right)} \xrightarrow[n \to +\infty]{} \prod_{k \in \mathbb{N}^*} \left(1 + \frac{x-1}{k}\right) e^{-\frac{x-1}{k}} \prod_{p \in \mathcal{P}} \frac{1 - \frac{1}{p}}{1 - \frac{x}{p}} e^{-\frac{x-1}{p}}$$

The interpretation of this last convergence is clear : one has the independent model

$$\widehat{\Omega}_n := \sum_{p \leqslant n} G_p$$

where $(G_p)_p$ is a sequence of independent random variables each of geometric distribution with parameter $1/p$ that correspond to the $(v_p(U_n))_p$, and a corrective factor which is the total number of cycles of a random uniform permutation.

A model that reproduces the mod-Poisson fluctuations of $\Omega(U_n)$ in the same vein as theorem 3.5.1 could be interesting to find and will be developed in the future.

Here, the delta-Zeta randomisation applies again and one finds naturally the mod-Poisson convergence of $\Omega(Z_\alpha) \stackrel{\mathcal{L}}{=} \sum_p \mathcal{G}(p^{-\alpha})$ when $\alpha \to 1$. The same reasonning as before allows to understand why this is the Γ factor that appears here again.

3.6.4 Perspectives and open problems

There are several directions to generalise the philosophy of additional randomisation in mod-* convergence.

One can try naturally to adapt the last results to mod-Gaussian convergence, in particular to the moments conjecture (since the speed of convergence is $\log\log T$ in (1.22), this amounts to look for a randomisation in T^Z for a certain random variable Z). Of course, one needs first to prove (or disprove) that the group factor Φ_U is the inverse of a Mellin transform.The mod-Lévy (and mod-Dickman) convergence can also be studied.

Another natural generalisation consists in staying in the mod-Poisson framework and changing the model, and in particular, to consider models where an additional randomisation appears naturally. This is the case for the cycle structure of a random uniform permutation, but also for polynomials over finite fields since the analogue of Polya's CIP is the cyclotomic identity (we use the notations of example 3.3.4 and denote by $c_k(P)$ the number of irreducible factors of degree k of the monic polynomial P, the sum being understood as a sum on monic polynomials)

$$\sum_{P \in \mathbb{F}_q[X]} t^{\deg P} \prod_{k \geqslant 1} z_k^{c_k(P)} = \prod_{\pi \in \mathcal{P}(\mathbb{F}_q[X])} \left(\frac{1}{1 - z_{\deg \pi}\, t^{\deg \pi}} \right)$$

or, equivalently, with $c_d^{(q)} := \operatorname{card}\{P \in \mathbb{F}_q[X] \ / \ \deg P = d\}$

$$\sum_{n \geqslant 0} c_n^{(q)} t^n \left(\frac{1}{c_n^{(q)}} \sum_{P \in \mathbb{F}_q[X],\, \deg P = n} \prod_{k \geqslant 1} z_k^{c_k(P)} \right) = \prod_{d \geqslant 1} \left(\frac{1}{1 - z_d t^d} \right)^{c_d^{(q)}}$$

Here, one must multiply this last identity by

$$\prod_{d \geqslant 1} \left(1 - t^d\right)^{c_d^{(q)}}$$

to get the probabilistic interpretation in terms of CIP and randomisation by the random variable G_t such that

$$\mathbb{P}\left(G_t = n\right) = t^n c_n^{(q)} \prod_{d \geqslant 1} \left(1 - t^d\right)^{c_d^{(q)}}$$

Of course, other divisibility structures can occur, in particular *logarithmic combinatorial structures* defined in [2], or other arithmetic structures such as $\mathbb{Z}[i]$, etc.

Nevertheless, the most challenging aspect of the randomisation would consist in finding a way to directly prove mod-* convergence of the non-randomised sequence using the randomised sequence and the independence structure (or any other structure) that arise from the randomisation. Indeed, the reasonning allowing to eliminate the corrective factor by means of a randomisation supposes that the initial sequence of random variables converges in the mod-* sense. The fact that there exists an independent model obtained by randomisation implies that if mod-* convergence exists, the only possible limiting function is the independent one corrected by a limiting function involving the fluctuations of the randomisation variable. But nothing guarantees the existence of mod-* convergence. To achieve this, the technique is derandomisation, which amounts to use the Selberg-Delange method for the delta-Zeta distribution. But other methods are still possible and necessitate further investigations.

Chapter 4

Stein's method and mod-* convergence

Let $(X_n)_n$ be a sequence of random variables converging in law to a variable $Z \sim \mathscr{N}(0,1)$, for instance, X_n is the sum of n i.i.d. random variables of expectation 0 and variance $1/\sqrt{n}$. The Central Limit Theorem asserts that

$$d_{\mathrm{Kol}}(X_n, Z) := \sup_{x\in\mathbb{R}} |\mathbb{P}(X_n \leqslant x) - \mathbb{P}(Z \leqslant x)| \xrightarrow[n\to+\infty]{} 0$$

The Berry-Esséen theorem ([17, 36]) is a direct continuation of the Central Limit Theorem : it gives the rate of convergence of this latest limit

$$d_{\mathrm{Kol}}(X_n, Z) \leqslant \frac{C}{\sqrt{n}}$$

where C is a constant depending on the sequence $(X_n)_n$.

The classical method to prove the Berry-Esséen theorem is to proceed to a Fourier inversion. This method applies perfectly in the framework of the sum of i.i.d. random variables thanks to the equality $\mathbb{E}\left(e^{-i\xi(X_1+X_2)}\right) = \mathbb{E}\left(e^{-i\xi X_1}\right)\mathbb{E}\left(e^{-i\xi X_2}\right)$, characteristic of independent variables. But in the case of a marked dependence, it becomes much more difficult to handle.

Stein's method was created by Charles Stein in [93] and allows to escape from the Fourier formalism to achieve the same goal, to prove the Berry-Esséen theorem. The key point of the method consists in using a characteristic operator to replace the characteristic function which is easier to handle in situations of dependency. Many paradigm shifts were then observed in the theory ; initially designed for the Gaussian distribution, the method was extended to the Poisson one [23] and the characteristic operator metamorphosed into a probabilistic transformation such as the 0-bias or the size-bias ones [44, 45], the characterisation of the distribution via the operator being then replaced by an equation in law using these transformations.

Stein's method can also be seen as a probabilistic tool to approximate expectations. This is the final form that Stein gave to his method in his Lecture Notes, *Approximate computations of expectations* [94]. Given a sequence of random variables $(Z_n)_n$ and a function f, writting that $\mathbb{E}\left(f(Z_n)\right) \approx \mathbb{E}\left(f(Z)\right)$ can also be understood as a convergence in law, and this is thus a way to prove such a convergence.

4.1 Reminder on metrics on probability spaces

For the sake of clarity, we will restrict ourselves to the case of measures on $\mathbb{R}$; the more general case of a Polish space (a complete separable metric space) is also possible.

4.1.1 The Kantorovich representation

As Stein's method is concerned with bounds in a probability metric, we remind some of the most classical ones. A typical distance in probability is of the form

$$d_{\mathcal{D}}(\mathbb{P}_1, \mathbb{P}_2) := \sup_{f \in \mathcal{D}} \left\{ \left| \int f d\mathbb{P}_1 - \int f d\mathbb{P}_2 \right| \right\}$$

with $\mathcal{D}$ a given space of functions.

We remark that in certain cases of spaces, such a metric can be used on general measures. Moreover, in optimal transport theory, this kind of metrics is said to have a *Kantorovich representation* (*cf.* [100]).

The total variation distance

Let $\mathscr{B}or(\mathbb{R})$ denote the space of Borel sets of $\mathbb{R}$. The total variation distance between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined to be

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) := \sup_{A \in \mathscr{B}or(\mathbb{R})} |\mathbb{P}_1(A) - \mathbb{P}_2(A)|$$

We thus have

$$\mathcal{D} = \{\mathbb{1}_A, A \in \mathscr{B}or(\mathbb{R})\}$$

We remark that $d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) \in [0, 1]$ for all probability measures $\mathbb{P}_1, \mathbb{P}_2$. It is clear that we can also define such a distance on arbitrary measurable spaces and not only on a metric space or a Polish space.

The Wasserstein distance

This is the distance W_1 defined by

$$W_1(\mathbb{P}_1, \mathbb{P}_2) := \sup_{f \in Lip(1)} \left| \int f d\mathbb{P}_1 - \int f d\mathbb{P}_2 \right|$$

where

$$\mathcal{D} = Lip(1) := \left\{ f : \mathbb{R} \to \mathbb{R} \ / \ \|f\|_{Lip} = 1 \right\}$$

the semi-norm $\|\cdot\|_{Lip}$ (that can take the value $+\infty$) being defined by

$$\|f\|_{Lip} := \sup_{x \neq y} \left| \frac{f(x) - f(y)}{x - y} \right|$$

We can define this metric on a more general metric space endowed with the distance d if we set $\|f\|_{Lip} := \sup_{x \neq y} \frac{|f(x)-f(y)|}{d(x,y)}$. We have $W_1(\mathbb{P}_1, \mathbb{P}_2) \in [0, +\infty]$.

Remark 4.1.1. In optimal transport theory, this metric is known as the *Kantorovich-Monge-Rubinstein* metric (see [100]).

Remark 4.1.2. As a lipschitz function is Lebesgue-almost everywhere differentiable, noting Df the function equal to the differential of f at the points where it is differentiable and $+\infty$ elsewhere, we have $\|f\|_{Lip} = \|Df\|_\infty$, where

$$\|f\|_\infty := \|f\|_{L^\infty} := \operatorname*{ess\,sup}_{x \in \mathbb{R}} |f(x)| = \inf \{a \in \mathbb{R} \ / \ \lambda(\{|f| > a\}) = 0\}$$

The Kolmogorov distance

This is the probability distance on $\mathbb{R}$ defined by

$$d_{\mathrm{Kol}}(\mathbb{P}_1, \mathbb{P}_2) := \sup_{x \in \mathbb{R}} |\mathbb{P}_1(] - \infty, x\,]) - \mathbb{P}_2(] - \infty, x\,])|$$

We thus have

$$\mathcal{D} = \left\{ \mathbb{1}_{]-\infty, x]}, x \in \mathbb{R} \right\}$$

This distance can be defined on $\mathbb{R}^n$ using cartesian products of sets of the form $] - \infty, x]$. For a general metric space, there is no definition.

As $\left\{ \mathbb{1}_{]-\infty, x]}, x \in \mathbb{R} \right\} \subset \mathscr{B}or(\mathbb{R})$, we have $d_{\mathrm{Kol}}(\mathbb{P}_1, \mathbb{P}_2) \leqslant d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2)$.

Remark 4.1.3. This distance is often called the "Kolmogorov-Smirnov" distance.

The Radon distance

This is the probability distance defined by

$$d_R(\mathbb{P}_1, \mathbb{P}_2) := \sup_{\|f\|_\infty \leqslant 1} \left| \int f d\mathbb{P}_1 - \int f d\mathbb{P}_2 \right|$$

where

$$\mathcal{D} = \mathcal{B}_\infty(0, 1) := \left\{ f \in \mathcal{C}^0(\mathbb{R}, \mathbb{R}) \ / \ \|f\|_\infty \leqslant 1 \right\}$$

This distance can be defined on arbitrary spaces considering the unit ball of a norm.

4.1.2 Link with the weak convergence

All the distances that were defined are stronger than the weak convergence, *i.e.* the convergence in law (or in distribution) seen as the convergence on the dual of the space of continuous function (or the weak-* convergence if you define the convergence on the space of probability measures). The convergence in law can be metrized by the Lévy-Prokhorov metric

$$d_{LP}(\mathbb{P}_1, \mathbb{P}_2) := \inf\left\{\varepsilon > 0 \ / \ \inf_{\mu \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)} \int \mathbb{1}_{\{|x-y|>\varepsilon\}} d\mu(x,y) \leqslant \varepsilon\right\}$$

where, noting $\mathcal{P}(\Omega)$ the space of probability distributions on a probability space Ω,

$$\Gamma(\mathbb{P}_1, \mathbb{P}_2) := \{\mu \in \mathcal{P}(\Omega \times \Omega) \ / \ \forall A \in \mathscr{B}or(\Omega), \quad \mu(A \times \Omega) = \mathbb{P}_1(A), \quad \mu(\Omega \times A) = \mathbb{P}_2(A)\}$$

Said differently, $\Gamma(\mathbb{P}_1, \mathbb{P}_2)$ is the space of *couplings* of $(\mathbb{P}_1, \mathbb{P}_2)$, that is, the probabilities on $\Omega \times \Omega$ that have marginals $\mathbb{P}_1$ and $\mathbb{P}_2$.

This interpretation of the weak convergence as a functional on couplings of the two measures can also be made on the latest probabilistic distances.

Remark 4.1.4. The distances presented are thus stronger than the weak convergence distance (that is : if a sequence of probability converges for these distances, it then converges in law). The opposite is not true : the distances are not equivalent.

4.1.3 The coupling form

With a slight abuse of notation, we write $(X, Y) \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)$ in place of $\gamma \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)$ if $X \sim \mathbb{P}_1$ and $Y \sim \mathbb{P}_2$. We thus have the coupling interpretation of all the latest distances :

- **Total variation distance :**

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\gamma \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)} \int \mathbb{1}_{\{x \neq y\}} d\gamma(x,y) = \inf_{(X,Y) \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{P}(X \neq Y)$$

- **Wasserstein distance :**

$$W_1(\mathbb{P}_1, \mathbb{P}_2) = \inf_{\gamma \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)} \int |x-y| \, d\gamma(x,y) = \inf_{(X,Y) \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}\left(|X - Y|\right)$$

This last distance can be extended into L^p Wasserstein distances defined by

$$W_p(\mathbb{P}_1, \mathbb{P}_2) := \inf_{\gamma \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)} \left(\int |x-y|^p \, d\gamma(x,y)\right)^{\frac{1}{p}} = \inf_{(X,Y) \in \Gamma(\mathbb{P}_1, \mathbb{P}_2)} \left(\mathbb{E}\left(|X-Y|^p\right)\right)^{\frac{1}{p}}$$

Remark 4.1.5. The total variation distance is a really strong notion of distance : if $(B_k)_k$ is a sequence of i.i.d. fair coins (Bernouilli random variables with values in $\{\pm 1\}$ and equal probability) and $Z \sim \mathscr{N}(0,1)$, setting $S_n := \sum_{k=1}^n B_k$, we have

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n \to \infty]{\mathcal{L}} \mathscr{N}(0,1), \qquad d_{\mathrm{Kol}}\left(\frac{S_n}{\sqrt{n}}, Z\right) \xrightarrow[n\to\infty]{} 0, \qquad W_1\left(\frac{S_n}{\sqrt{n}}, Z\right) \xrightarrow[n\to\infty]{} 0$$

but $\forall\, n \geqslant 1$

$$d_{\mathrm{TV}}\left(\frac{S_n}{\sqrt{n}}, Z\right) = 1$$

4.2 The first form of Stein's method for the normal approximation

We introduce the Stein's method for the normal approximation, and will be concerned with the case of a more general distribution later.

4.2.1 The Stein's operator

The main idea of Stein's method is to replace a *global problem* (approximating a distribution by another one showing that their characteristic functions are close) by a *local problem* (approximating a distribution by another one showing that certain characteristic operators are close). We thus replace the characteristic function by the *characteristic operator*.

Definition 4.2.1 (Characteristic operator)**.** Let $\mathcal{H}$ be a given space of real functions and $\mathcal{A}$ an operator acting on $\mathcal{H}$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $X, Y : \Omega \to \mathbb{R}$ two real random variables. We denote by $\mathbb{P}_X$ the image measure of $\mathbb{P}$ by X, that is : $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$ for all $A \in \mathcal{F}$.

We say that $\mathcal{A}$ is a characteristic operator of the distribution $\mathbb{P}_X$ if for all Y

$$\mathbb{P}_Y = \mathbb{P}_X \Longleftrightarrow \mathbb{E}\left(\mathcal{A}h(Y)\right) = 0 \quad \forall\, h \in \mathcal{H}$$

Defined like this, the characteristic operator is an incomplete notion : one should always consider the couple $(\mathcal{A}, \mathcal{H})$. We nevertheless commit this abuse of concept and will define the relevant space $\mathcal{H}$ everytime needed.

As an application, we give a characteristic operator of the Gaussian distribution $\mathscr{N}(0,1)$.

Lemma 4.2.2 (Stein's lemma)**.** *Let $\mathcal{A}$ the operator defined by*

$$\mathcal{A}h(x) := h'(x) - xh(x) \tag{4.1}$$

defined on the space

$$\mathcal{H} := \left\{ f \in \mathcal{C}^1(\mathbb{R},\mathbb{R}) \cap \mathcal{C}_b^0(\mathbb{R},\mathbb{R}) \ / \ f' \in L^1(\mathbb{P}_G) \right\} \tag{4.2}$$

with $G \sim \mathscr{N}(0,1)$ and where $\mathcal{C}^1$ can be replaced by the absolutely continuous functions and $\mathcal{C}_b^0$ denotes the space of continuous bounded functions.

Then : $\mathcal{A}$ is a characteristic operator of $\mathscr{N}(0,1)$, i.e.

$$Z \sim \mathscr{N}(0,1) \Longleftrightarrow \mathbb{E}\left(\mathcal{A}h(Z)\right) = 0 \quad \forall\, h \in \mathcal{H}$$

Proof. Let $Z \sim \mathscr{N}(0,1)$, and let us suppose for the moment that $\mathcal{A}$ acts on $L^2(\mathbb{R})$ (usual L^2 space for the Lebesgue measure), *i.e.* that $\mathcal{A}h \in L^2(\mathbb{R})$, and that $\mathcal{H}$ is composed with functions in $\mathcal{C}_0^1$ that go to 0 in $\pm\infty$. We then have

$$\begin{aligned}
\mathbb{E}\left(\mathcal{A}h(Z)\right) &= \int_{\mathbb{R}} \mathcal{A}h(x) f_Z(x)dx \qquad \text{with } f_Z(x) = \tfrac{e^{-x^2/2}}{\sqrt{2\pi}} \\
&= \langle \mathcal{A}h, f_Z \rangle_{L^2} \qquad \text{if we suppose } \mathcal{A}h \in L^2(\mathbb{R}) \\
&= \langle h, \mathcal{A}^* f_Z \rangle_{L^2}
\end{aligned}$$

We have denoted by $\mathcal{A}^*$ the L^2-adjoint of the operator $\mathcal{A}$ that we suppose acting on L^2 for the moment. Thus, we have a characteristic operator if and only if $\mathcal{A}^* f_Z = 0$. By integration by parts, we have

$$D^* = -D$$

with $D : f \mapsto f'$ the operator of differentiation defined on $L^2(\mathbb{R}) \cap \mathcal{H}$. Thus, by the usual rules of adjointness (such as $(AB)^* = B^* A^*$ for operators A, B and $\mathfrak{M}_f^* = \mathfrak{M}_f$ if $\mathfrak{M}_f : g \mapsto fg$), we have

$$\mathcal{A}^* f(x) = -f'(x) - x f(x)$$

It is clear that $f_Z'(x) = x f_Z(x)$, *i.e.* that $\mathcal{A}^* f_Z = 0$.

In the general case, with the space $\mathcal{H}$ defined by (4.2), we show that $\int_{\mathbb{R}} \mathcal{A}h(x) f_Z(x) dx = 0$ by an integration by parts, and reciprocally, if X is such that $\int_{\mathbb{R}} \mathcal{A}h(x) f_X(x) dx = 0$, by specializing h to the functions $h_\lambda : x \mapsto e^{i\lambda x}$, we prove the result.

□

Remark 4.2.3. This method to obtain a Stein operator is general. When the distribution is discrete, we replace the space $L^2(\mathbb{R})$ by the space $\ell^2(\Omega)$. For instance, for the Poisson distribution, $\Omega = \mathbb{N}$ and if $X \sim \mathscr{P}(1)$, $\mathbb{P}(X = k) = e^{-1} \frac{1}{k!} =: p_k$. We have $p_k = \frac{1}{ek \cdot (k-1)!} = \frac{p_{k-1}}{k}$ so the operator $\mathcal{A}^*$ defined by $\mathcal{A}^* f(k) := k f(k) - f(k-1)$ cancels $(p_k)_k$. We need to take its adjoint for the usual scalar product on $\ell^2(\mathbb{N})$. Denote by θ the shift operator $\theta f(k) := f(k+1)$. We have, for $f, g \in \ell^2(\mathbb{N})$ with the convention that $g(-1) = 0$,

$$\langle \theta f, g \rangle_{\ell^2} = \sum_{k \geqslant 0} f(k+1) g(k) = \sum_{k \geqslant 1} f(k) g(k-1) = \langle f, \theta^{-1} g \rangle_{\ell^2} \qquad \text{i.e. } \theta^* = \theta^{-1}$$

so we must take $\mathcal{H} = \left\{ f \in \ell^2(\mathbb{N}) \ / \ f(0) = 0 \right\}$ and $\mathcal{A} f(k) := k f(k) - f(k+1)$ to obtain

$$Z \sim \mathscr{P}(1) \Longleftrightarrow \mathbb{E}\left(Z h(Z)\right) = \mathbb{E}\left(h(Z+1)\right) \quad \forall\, h \in \mathcal{H}$$

We can moreover weaken the hypothesis on the integrability and only suppose that $\mathcal{H} = \{f \ / \ \mathbb{E}\left(|Zh(Z)|\right) < \infty, \quad f(0) = 0\}$.

How can we use such a characterisation to get bounds in probabilistic metrics ? The methodology consists in solving the so-called *Stein's equation*, namely to invert the operator $\mathcal{A}$ on its domain of invertibility (which amounts to take the *pseudo-inverse*).

$$\mathcal{A} f = g \Longleftrightarrow f = \mathcal{A}^{-1} g$$

For such a goal, we remark that taking the expection of $\mathcal{A} f(Z) = g(Z)$ with $Z \sim \mathscr{N}(0,1)$ implies that $\mathbb{E}\left(g(Z)\right) = 0$, which imposes that g is of the form $h - \mathbb{E}\left(h(Z)\right)$. This gives the domain of invertibility. We thus solve the equation

$$\mathcal{A} f = h - \mathbb{E}\left(h(Z)\right) \Longleftrightarrow f = \mathcal{A}^{-1}\left(h - \mathbb{E}\left(h(Z)\right)\right)$$

Now, write

$$\begin{aligned}\Phi(h) &:= \mathbb{E}\left(h(Z)\right) \quad \text{ for } Z \sim \mathscr{N}(0,1)\\ h_\Phi(x) &:= h - \Phi(h)\end{aligned}$$

The solution of the equation

$$\mathcal{A}f = h_\Phi \tag{4.3}$$

is given by

$$\mathcal{A}^{-1}h_\Phi(x) = e^{x^2/2}\int_{-\infty}^{x} h_\Phi(t)e^{-t^2/2}dt = \frac{\mathbb{E}\left(h_\Phi(Z)\mathbb{1}_{\{Z\leqslant x\}}\right)}{f_Z(x)}$$

As $\mathbb{E}\left(h_\Phi(Z)\right) = \mathbb{E}\left(h(Z) - \mathbb{E}\left(h(Z)\right)\right) = 0$, we also have

$$\mathcal{A}^{-1}h_\Phi(x) = \frac{\mathbb{E}\left(h_\Phi(Z)\left(1-\mathbb{1}_{\{Z>x\}}\right)\right)}{f_Z(x)} = -\frac{\mathbb{E}\left(h_\Phi(Z)\mathbb{1}_{\{Z\geqslant x\}}\right)}{f_Z(x)}$$

Remark 4.2.4. We can express the solution of the Stein's equation in terms of Malliavin calculus (*cf.* [79]) and it then takes the form of a conditional expectation

$$\mathcal{A}^{-1}\left(h - \mathbb{E}\left(h(Z)\right)\right)(x) = \mathbb{E}\left(\varphi_Z(h)|Z=x\right)$$

where $\varphi_Z(h)$ can be given in terms of Malliavin calculus[1].

The Stein's methodology consists in writing

$$\mathbb{E}\left(h(X)\right) - \mathbb{E}\left(h(Z)\right) = \mathbb{E}\left(h(X) - \mathbb{E}\left(h(Z)\right)\right) = \mathbb{E}\left(h_\Phi(X)\right) = \mathbb{E}\left(\mathcal{A}\mathcal{A}^{-1}h_\Phi(X)\right)$$

so that we can bound the distance

$$d_{\mathcal{H}}(X,Y) = \sup_{h\in\mathcal{H}}\left|\mathbb{E}\left(h(X)\right) - \mathbb{E}\left(h(Z)\right)\right| \tag{4.4}$$

For $h \in \mathcal{H}$, we define

$$f_h := \mathcal{A}^{-1}h_\Phi \tag{4.5}$$

so that f_h solves the Stein's equation

$$\mathcal{A}f_h(x) := f_h'(x) - xf_h(x) = h_\Phi(x) := h(x) - \Phi(h)$$

We thus have

$$d_{\mathcal{H}}(X,Z) = \sup_{h\in\mathcal{H}}\left|\mathbb{E}\left(\mathcal{A}f_h(X)\right)\right| = \sup_{h\in\mathcal{H}}\left|\mathbb{E}\left(f_h'(X) - Xf_h(X)\right)\right| \tag{4.6}$$

The idea is thus to bound the right side of (4.6) hoping that the properties of f_h and the structure of X will give a sufficient bound.

[1] More precisely, if $f(Z)$ is in the Wiener space $\mathbb{D}^{1,2}$, we have $\varphi_Z(h) = \langle DL^{-1}f(Z), DZ\rangle_{\mathfrak{H}}$ where D denotes the Malliavin derivative, L is the infinitesimal generator of the Ornstein-Uhlenbeck process and $\langle\cdot,\cdot\rangle_{\mathfrak{H}}$ denotes the scalar product on the Wiener space.

4.2.2 Some estimates

The boundedness properties of $f_h := \mathcal{A}^{-1}h_\Phi$ are a corollary of the properties of the operator $\mathcal{A}$. In the case of the Gaussian distribution, $\mathcal{A}$ is a first order differential operator so its pseudo-inverse $\mathcal{A}^{-1}$ is a first order integral operator, whose operator norm $\left\|\mathcal{A}^{-1}\right\|_{L^\infty \to L^\infty}$ is bounded. But we must first specify the space of functions that we choose, that is, the probability distance that we choose : the behavior of f_h depends drastically on the properties of h.

Lemma 4.2.5. *Let $f_h := \mathcal{A}^{-1}h_\Phi$ the solution of the Stein's equation with $h \in \mathcal{H}$ to be specified. Then :*

1. *If h is bounded ($\|h\|_\infty < \infty$), and D denotes the operator of differentiation, then*

$$\begin{aligned} \left\|\mathcal{A}^{-1}h_\Phi\right\|_\infty &\leqslant \sqrt{\frac{\pi}{2}}\,\|h_\Phi\|_\infty \\ \left\|D\mathcal{A}^{-1}h_\Phi\right\|_\infty &\leqslant 2\;\;\|h_\Phi\|_\infty \end{aligned} \tag{4.7}$$

2. *If h is absolutely continuous ($\|Dh\|_\infty < \infty$), then*

$$\begin{aligned} \left\|\mathcal{A}^{-1}h_\Phi\right\|_\infty &\leqslant 2\,\|Dh\|_\infty \\ \left\|D\mathcal{A}^{-1}h_\Phi\right\|_\infty &\leqslant \sqrt{\frac{\pi}{2}}\,\|Dh\|_\infty \\ \left\|D^2\mathcal{A}^{-1}h_\Phi\right\|_\infty &\leqslant 2\,\|Dh\|_\infty \end{aligned} \tag{4.8}$$

Proof.

1. • For $x > 0$, we have

$$\mathcal{A}^{-1}h_\Phi(x) = -\frac{\mathbb{E}\left(h_\Phi(Z)\mathbb{1}_{\{Z \geqslant x\}}\right)}{f_Z(x)}$$

so that

$$\begin{aligned} \left|\mathcal{A}^{-1}h_\Phi(x)\right| &\leqslant \frac{\mathbb{E}\left(\left|h_\Phi(Z)\right|\mathbb{1}_{\{Z \geqslant x\}}\right)}{f_Z(x)} \\ &\leqslant \|h_\Phi\|_\infty \frac{\mathbb{E}\left(\mathbb{1}_{\{Z \geqslant x\}}\right)}{f_Z(x)} \\ &\leqslant \|h_\Phi\|_\infty \sup_{x>0} \frac{\mathbb{P}\left(Z \geqslant x\right)}{f_Z(x)} \\ &= \frac{1/2}{1/\sqrt{2\pi}}\|h_\Phi\|_\infty = \sqrt{\frac{\pi}{2}}\,\|h_\Phi\|_\infty \end{aligned}$$

The fact that $x \mapsto \mathbb{P}\left(Z \geqslant x\right)/f_Z(x)$ is decreasing on $\mathbb{R}_+$ (hence reaches its maximal value in 0) comes from the fact that

$$\frac{d}{dx}\frac{\mathbb{P}\left(Z \geqslant x\right)}{f_Z(x)} = x\frac{\mathbb{P}\left(Z \geqslant x\right)}{f_Z(x)} - 1$$

and the familiar fact that for $x > 0$

$$\mathbb{P}(Z \geqslant x) = \int_x^{+\infty} e^{-u^2/2} du < \int_x^{+\infty} \frac{x}{u} e^{-u^2/2} du = \frac{e^{-x^2/2}}{x\sqrt{2\pi}} = \frac{f_Z(x)}{x} \tag{4.9}$$

We proceed analogously for $x < 0$ using $\mathcal{A}^{-1} h_\Phi(x) = \frac{\mathbb{E}\left(h_\Phi(Z) \mathbb{1}_{\{Z \leqslant x\}}\right)}{f_Z(x)}$ and the inequality for $x < 0$

$$\mathbb{P}(Z \leqslant x) < \frac{e^{-x^2/2}}{|x| \sqrt{2\pi}}$$

- We have $D\mathcal{A}^{-1} h_\Phi(x) = h_\Phi(x) + x\mathcal{A}^{-1} h_\Phi(x)$ by definition of $\mathcal{A}$. Thus, for $x \geqslant 0$

$$\begin{aligned}
\left|D\mathcal{A}^{-1} h_\Phi(x)\right| &\leqslant |h_\Phi(x)| + |x| \left|\mathcal{A}^{-1} h_\Phi(x)\right| \\
&\leqslant \|h_\Phi\|_\infty \left(1 + \sup_{x>0} \frac{|x| \, \mathbb{E}\left(\mathbb{1}_{\{Z \geqslant x\}}\right)}{f_Z(x)}\right) \\
&\leqslant 2 \|h_\Phi\|_\infty
\end{aligned}$$

the last equality coming from (4.9). We proceed analogously for $x \leqslant 0$.

2. The proof of this part is left as an exercise.

□

Now that we have such estimates, we can choose the type of function (bounded or absolutely continuous) corresponding to a choice of norm (Kolmogorov or Wasserstein) to estimate the distance between two distributions. For differentiability reasons, it appears that it is simpler to work in the Wasserstein metric.

4.2.3 An example : the Central Limit Theorem

Stein's method allows to prove the following theorem that will imply the central limit theorem.

Theorem 4.2.6. *Let $(X_k)_k$ a sequence of independent random variables with $\mathbb{E}(X_k) = 0$, $\mathbb{E}\left(X_k^2\right) = 1$, and $\mathbb{E}\left(|X_k|^3\right) < \infty$. Let $Z \sim \mathcal{N}(0,1)$. Then, we have the following **Berry-Esséen bound in the Wasserstein metric***

$$W_1\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k, Z\right) \leqslant \frac{3}{n^{3/2}} \sum_{k=1}^n \mathbb{E}\left(|X_k|^3\right)$$

Proof. We set

$$\begin{aligned} S_n &:= \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \\ S_n^{(k)} &:= S_n - \frac{1}{\sqrt{n}}X_k = \frac{1}{\sqrt{n}}\sum_{i\neq k} X_i \end{aligned}$$

Let h such that $\|h'\|_\infty < \infty$ and $\|h''\|_\infty < \infty$; let $f := f_h = \mathcal{A}^{-1}h_\Phi$ be the solution of the Stein equation (4.3).

The first step of the method is to write

$$\begin{aligned} \mathbb{E}\left(\mathcal{A}\mathcal{A}^{-1}h(S_n)\right) &= \mathbb{E}\left(S_n f(S_n) - f'(S_n)\right) \\ &= \mathbb{E}\left(S_n f(S_n) - \frac{1}{n}\sum_{k=1}^n f'\left(S_n^{(k)}\right)\right) + \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^n f'\left(S_n^{(k)}\right) - f'(S_n)\right) \\ &= \sum_{k=1}^n \mathbb{E}\left(\frac{X_k}{\sqrt{n}} f(S_n) - \frac{1}{n} f'\left(S_n^{(k)}\right)\right) + \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left(f'\left(S_n^{(k)}\right) - f'(S_n)\right) \end{aligned}$$

Although the introduction of the term $\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^n f'\left(S_n^{(k)}\right)\right)$ seems magic, it will be motivated in the chapter 4.3. Thus

$$\begin{aligned} \left|\mathbb{E}\left(\mathcal{A}\mathcal{A}^{-1}h(S_n)\right)\right| &\leqslant \sum_{k=1}^n \left|\mathbb{E}\left(\frac{X_k}{\sqrt{n}} f(S_n) - \frac{1}{n} f'\left(S_n^{(k)}\right)\right)\right| + \frac{1}{n}\sum_{k=1}^n \left|\mathbb{E}\left(f'\left(S_n^{(k)}\right) - f'(S_n)\right)\right| \\ &=: J_1 + J_2 \end{aligned}$$

We now deal with these two terms :

- For J_2, we write

$$\begin{aligned} J_2 &\leqslant \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left(\left|f'\left(S_n^{(k)}\right) - f'(S_n)\right|\right) \\ &\leqslant \|f''\|_\infty \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left(\left|S_n^{(k)} - S_n\right|\right) \\ &= \frac{\|f''\|_\infty}{n\sqrt{n}}\sum_{k=1}^n \mathbb{E}\left(|X_k|\right) \end{aligned}$$

  We remark that with the assumptions,

$$\mathbb{E}\left(|X_k|^3\right) \geqslant \left[\mathbb{E}\left(|X_k|^2\right)\right]^{3/2} = 1 \quad \Longrightarrow \quad \mathbb{E}\left(|X_k|\right) \leqslant \left[\mathbb{E}\left(|X_k|^3\right)\right]^{1/3} \leqslant \mathbb{E}\left(|X_k|^3\right)$$

  Thus,

$$J_2 \leqslant \frac{\|f''\|_\infty}{n^{3/2}}\sum_{k=1}^n \mathbb{E}\left(|X_k|^3\right)$$

- For J_1, we use the

  Lemma 4.2.7. *We have the bound*

$$\left|\mathbb{E}\left(\frac{X_k}{\sqrt{n}}f(S_n)-\frac{1}{n}f'\left(S_n^{(k)}\right)\right)\right|\leqslant\frac{1}{2}\frac{\|f''\|_\infty}{n^{3/2}}\mathbb{E}\left(|X_k|^3\right)$$

  Indeed, write the Taylor expansion

$$f(S_n)=f\left(S_n^{(k)}\right)+\left(S_n-S_n^{(k)}\right)f'\left(S_n^{(k)}\right)+\int_{S_n^{(k)}}^{S_n}(S_n-t)f''(t)dt$$

  Thus

$$\mathbb{E}\left(\frac{X_k}{\sqrt{n}}\left[f(S_n)-f\left(S_n^{(k)}\right)-\left(S_n-S_n^{(k)}\right)f'\left(S_n^{(k)}\right)\right]\right)=\mathbb{E}\left(\frac{X_k}{\sqrt{n}}\int_{S_n^{(k)}}^{S_n}(S_n-t)f''(t)dt\right)$$

  Take the absolute value of each side. The RHS becomes

$$\begin{aligned}\left|\mathbb{E}\left(\frac{X_k}{\sqrt{n}}\int_{S_n^{(k)}}^{S_n}(S_n-t)f''(t)dt\right)\right| &\leqslant \mathbb{E}\left(\left|\frac{X_k}{\sqrt{n}}\int_{S_n^{(k)}}^{S_n}(S_n-t)f''(t)dt\right|\right)\\ &\leqslant \|f''\|_\infty\ \mathbb{E}\left(\frac{|X_k|}{\sqrt{n}}\left|\int_{S_n^{(k)}}^{S_n}|S_n-t|\,dt\right|\right)\\ &= \|f''\|_\infty\ \mathbb{E}\left(\frac{|X_k|}{\sqrt{n}}\frac{1}{2}\left(S_n^{(k)}-S_n\right)^2\right)\\ &= \frac{1}{2}\|f''\|_\infty\ \mathbb{E}\left(\left(\frac{|X_k|}{\sqrt{n}}\right)^3\right)=\frac{\|f''\|_\infty}{2n^{3/2}}\ \mathbb{E}\left(|X_k|^3\right)\end{aligned}$$

  The LHS is such that

  - $\mathbb{E}\left(X_kf\left(S_n^{(k)}\right)\right)=\mathbb{E}\left(X_k\right)\mathbb{E}\left(f\left(S_n^{(k)}\right)\right)=0$ by independance and as $\mathbb{E}\left(X_k\right)=0$,
  - $\mathbb{E}\left(X_k^2f'\left(S_n^{(k)}\right)\right)=\mathbb{E}\left(X_k^2\right)\mathbb{E}\left(f'\left(S_n^{(k)}\right)\right)=\mathbb{E}\left(f'\left(S_n^{(k)}\right)\right)$ as $\mathbb{E}\left(X_k^2\right)=1$.

  Thus, the LHS is equal to

$$\mathbb{E}\left(\frac{X_k}{\sqrt{n}}f(S_n)-\frac{1}{n}f'\left(S_n^{(k)}\right)\right)$$

  which proves the lemma.

Now, we can finally prove the W_1-Berry-Esséen bound. We have

$$\left|\mathbb{E}\left(S_n f(S_n) - f'(S_n)\right)\right| \leqslant J_1 + J_2 = \left(1 + \frac{1}{2}\right)\frac{1}{n^{3/2}}\sum_{k=1}^{n}\mathbb{E}\left(|X_k|^3\right)\|f''\|_\infty$$

Using the bounds (4.46), we have

$$\left\|D^2\mathcal{A}^{-1}h_\Phi\right\|_\infty = \left\|f_h''\right\|_\infty \leqslant 2\|Dh\|_\infty$$

Thus, taking the supremum over $\{h \ / \ \|Dh\|_\infty \leqslant 1\}$, we get the L^1-Wasserstein distance, which gives the result :

$$W_1\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k, Z\right) \leqslant \frac{3}{n^{3/2}}\sum_{k=1}^{n}\mathbb{E}\left(|X_k|^3\right)$$

□

Remark 4.2.8. The real Berry-Esséen bound uses the Kolmogorov distance and not the Wasserstein distance. But we have

$$d_{\mathrm{Kol}}(Y, Z) \leqslant \frac{2}{\sqrt[4]{2\pi}}\sqrt{W_1(Y, Z)}$$

Unfortunately, this gives a speed of convergence of order $n^{-1/4}$, which is suboptimal. A true Berry-Esséen bound would require a sharp analysis of the solution of Stein's equation for functions h of the form $h(x) = \mathbb{1}_{\{x \leqslant t\}}$. See [94] for details.

4.3 Stein's method with exchangeable pair

The introduction of the factor $\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n} f'\left(S_n^{(k)}\right)\right)$ in the last proof can be explained by a special symmetry property of the sequence $(S_n)_n$ whose ultimate formalism allows to use the Stein's method automatically once such a symmetry is discovered.

Definition 4.3.1 (Stein pair). We say that a pair of random variables (W, W') is an **exchangeable pair** if

$$(W, W') \overset{\mathcal{L}}{=} (W', W)$$

We say that this is an a-**Stein pair** if it moreover satisfies :

$$\mathbb{E}\left(W - W' | W\right) = aW \tag{4.10}$$

We thus have $\mathbb{E}(W'|W) = (1-a)W$ wich informally means that W' is almost (up to some constant a) equal to W in the sense of the projection. This is the idea that you have only modified W with a *small perturbation* and that this particular coupling reduces the range of possible distances (for instance, for distances that can be expressed by means of a coupling).

Example 4.3.2. Let $I \sim \mathcal{U}(\llbracket 1,n \rrbracket)$ a uniform random variable in the set $\llbracket 1,n \rrbracket$, and consider $S_n := \sum_{k=1}^n X_k$. Let $(X'_k)_k \sim iid(X)$ such that $(X'_k)_k$ is independent of $(X_k)_k$. We also suppose that I is independent of $(X_k)_k$ and $(X'_k)_k$. Define

$$S_n^{(I)} := S_n - X_I + X'_I = \sum_{k \neq I} X_k + X'_I$$

Then, $\left(S_n, S_n^{(I)}\right)$ is a $1/n$-Stein exchangeable pair. It is clear that $\left(S_n, S_n^{(I)}\right) \overset{\mathcal{L}}{=} \left(S_n^{(I)}, S_n\right)$ since we have just replaced X_I by X'_I in S_n and that $X'_I \overset{\mathcal{L}}{=} X_I$ in addition to being independent of all the other terms of the sequence. Moreover

$$\begin{aligned}\mathbb{E}\left(S_n - S_n^{(I)} \middle| X_1, \ldots, X_n\right) &= \mathbb{E}\left(X_I - X'_I | X_1, \ldots, X_n\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(X_i - X'_i | X_1, \ldots, X_n\right) \\ &= \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{n} S_n\end{aligned}$$

This implies, by conditionning on $\sigma(S_n) \subset \sigma(X_1, \ldots, X_n)$ that

$$\mathbb{E}\left(S_n - S_n^{(I)} \middle| S_n\right) = \frac{1}{n} S_n$$

The philosophy of the exchangeable pair is that, when encoutered, the problem is considerably simplified. In particular, it allows to use an abstract approximation theorem that can be applied anytime needed to a wide range of situations.

Remark 4.3.3. When such an exchangeable pair is not available but one has an approximation with rest of the form

$$\mathbb{E}\left(W'|W\right) = (1-a)W + \varepsilon(W)$$

with $\varepsilon(W)$ controlled, the last philosophy also applies (see e.g. [94]).

Lemma 4.3.4 (Polarisation of an antisymmetric functionnal using an exchangeable pair)**.** *Let (W, W') an a-Stein pair and f a function such that $\mathbb{E}(|Wf(W)|) < \infty$. Then, setting $\Delta W := W' - W$, we get :*

$$\mathbb{E}(Wf(W)) = \frac{1}{2a}\mathbb{E}(\Delta W \Delta f(W)) \tag{4.11}$$

Proof.

$$\begin{aligned}\mathbb{E}\left(\Delta W \Delta f(W)\right) &= \mathbb{E}\left(f(W)\Delta W\right) - \mathbb{E}\left(f(W')\Delta W\right)\\ &= 2\mathbb{E}\left(f(W)\Delta W\right) \qquad \text{by exchangeability}\\ &= 2\mathbb{E}\left(f(W)\mathbb{E}\left(\Delta W | W\right)\right) = 2a\mathbb{E}\left(W f(W)\right).\end{aligned}$$

□

We can now state the *abstract approximation theorem* in the Gaussian context :

Theorem 4.3.5 (Stein's bounds on $\mathbb{E}\left(\mathcal{A}f(W)\right)$). *If $f \in \mathcal{C}^2_{0,0} := \left\{f \in \mathcal{C}^2 / \left\|f'\right\|_\infty, \left\|f''\right\|_\infty < \infty\right\}$, then*

$$\left|\mathbb{E}\left(\mathcal{A}f(W)\right)\right| \leqslant \left\|f'\right\|_\infty \left\|1 - \frac{1}{2a}\mathbb{E}\left(\left(\Delta W\right)^2 \middle| W\right)\right\|_1 + \left\|f''\right\|_\infty \frac{1}{4a}\left\|\Delta W\right\|_3^3 \tag{4.12}$$

Proof. Let us set $\underline{\Delta} = -\Delta$

$$\begin{aligned}\mathbb{E}\left(\mathcal{A}f(W)\right) &= \mathbb{E}\left(f'(W) - W f(W)\right) = \mathbb{E}\left(f'(W) - \frac{1}{2a}\Delta W \Delta f(W)\right) \qquad \text{by (4.11)}\\ &= \mathbb{E}\left(f'(W) - f'(W)\frac{1}{2a}\left(\Delta W\right)^2 + f'(W)\frac{1}{2a}\left(\Delta W\right)^2 - \frac{1}{2a}\Delta W \Delta f(W)\right)\\ &= \mathbb{E}\left(f'(W)\left[1 - \frac{1}{2a}\left(\Delta W\right)^2\right] - \frac{1}{2a}\underline{\Delta} W\left[\underline{\Delta} f(W) - f'(W)\underline{\Delta} W\right]\right)\end{aligned}$$

We thus have :

$$\left|\mathbb{E}\left(\mathcal{A}f(W)\right)\right| \leqslant \mathbb{E}\left(\left|f'(W)\left[1 - \frac{1}{2a}\mathbb{E}\left(\left(\Delta W\right)^2 \middle| W\right)\right]\right|\right) + \frac{1}{2a}\mathbb{E}\left(\left|\underline{\Delta} W\left[\underline{\Delta} f(W) - f'(W)\underline{\Delta} W\right]\right|\right)$$

The first term is such that :

$$\begin{aligned}\mathbb{E}\left(\left|f'(W)\left[1 - \frac{1}{2a}\mathbb{E}\left(\left(\Delta W\right)^2 \middle| W\right)\right]\right|\right) &\leqslant \left\|f'\right\|_\infty \mathbb{E}\left(\left|1 - \frac{1}{2a}\mathbb{E}\left(\left(\Delta W\right)^2 \middle| W\right)\right|\right)\\ &= \left\|f'\right\|_\infty \left\|1 - \frac{1}{2a}\mathbb{E}\left(\left(\Delta W\right)^2 \middle| W\right)\right\|_1\end{aligned}$$

For the second term, we distinguish two cases. If $W < W'$, we have :

$$\begin{aligned}\underline{\Delta} f(W) - f'(W)\underline{\Delta} W = f(W') - f(W) - (W' - W)f'(W) &= \int_W^{W'} \left(f'(u) - f'(W)\right) du\\ &= \int_W^{W'}\int_W^u f''(v)dvdu = \int_W^{W'}\int_W^{W'} \mathbb{1}_{\{v<u\}} f''(v)dvdu\\ &= \int_W^{W'} (W' - v) f''(v)dv\end{aligned}$$

Hence,

$$\left|\underline{\Delta} f(W) - f'(W)\underline{\Delta} W\right| \leqslant \|f''\|_\infty \int_W^{W'} |W' - v|\, dv = \|f''\|_\infty \frac{(\Delta W)^2}{2}$$

And we finally get :

$$\frac{1}{2a}\mathbb{E}\left(\left|\underline{\Delta} W\left[\underline{\Delta} f(W) - f'(W)\underline{\Delta} W\right]\right| \mathbb{1}_{\{W<W'\}}\right) \leqslant \frac{1}{4a}\|f''\|_\infty \mathbb{E}\left(|\Delta W|^3 \mathbb{1}_{\{W<W'\}}\right)$$

If $W' < W$, it is not hard to see that we get exactly the same bound. Thus, summing, we finally get the desired bound.

□

Remark 4.3.6. In the precedent result, we can notice that nothing is specified about a sequence of random variables converging in law. Nevertheless, this is the Stein's operator of the Gaussian distribution which is involved, so this bound is only useful for Gaussian approximation.

Last step of the approximation, we can now use the bounds (4.44) and (4.46)

$$\begin{aligned}
\left\|\mathcal{A}^{-1}\left(h - \mathbb{E}\left(h(Z)\right)\right)\right\|_\infty &\leqslant 2\left\|h - \mathbb{E}\left(h(Z)\right)\right\|_\infty \\
\left\|D\mathcal{A}^{-1}\left(h - \mathbb{E}\left(h(Z)\right)\right)\right\|_\infty &\leqslant \sqrt{\frac{\pi}{2}}\left\|h - \mathbb{E}\left(h(Z)\right)\right\|_\infty \\
\left\|D^2\mathcal{A}^{-1}\left(h - \mathbb{E}\left(h(Z)\right)\right)\right\|_\infty &\leqslant 2\left\|h'\right\|_\infty
\end{aligned}$$

to get the Stein's bounds :

$$\left|\mathbb{E}\left(h(W)\right) - \mathbb{E}\left(h(Z)\right)\right| \leqslant \sqrt{\frac{\pi}{2}}\left\|h - \mathbb{E}\left(h(Z)\right)\right\|_\infty \left\|1 - \frac{1}{2a}\mathbb{E}\left((\Delta W)^2 \,\middle|\, W\right)\right\|_1 + \left\|h'\right\|_\infty \frac{\|\Delta W\|_3^3}{2a} \tag{4.13}$$

As a corollary, we can apply this bound to the sum of independent random variables and recover the W_1-Berry-Esséen bound.

Corollary 4.3.7. *Let $(X_k)_{k\leqslant n}$ be independent real random variables such that $\mathbb{E}(X_k) = 0$, $\mathbb{E}\left(X_k^2\right) = \sigma_k^2$, $\mathbb{E}\left(X_k^4\right) < \infty$ and*

$$\sum_{k=1}^n \sigma_k^2 = 1$$

We set $W_n := \sum_{k=1}^n X_k$.
Then, for all $h \in \mathcal{C}_b^0(\mathbb{R}) \cap \mathcal{C}_b^1(\mathbb{R})$ (i.e. with $\|h\|_\infty < \infty$ and $\|h'\|_\infty < \infty$)

$$\left|\mathbb{E}\left(h(W)\right) - \mathbb{E}\left(h(Z)\right)\right| \leqslant \left\|h - \mathbb{E}\left(h(Z)\right)\right\|_\infty \sqrt{\frac{\pi}{2}\sum_{k=1}^n \left(\mathbb{E}\left(X_k^4\right) - \sigma_k^4\right)} + \left\|h'\right\|_\infty \frac{1}{2}\sum_{k=1}^n \left(\mathbb{E}\left(X_k^3\right) + 3\sigma_k^3\right)$$

Proof. Let $X' \overset{\mathcal{L}}{=} X$ be independent random variables of expectation 0 and variance σ^2 and finite fourth moment. Then

$$\begin{aligned}\mathbb{E}\left(|X-X'|^3\right) &\leqslant \mathbb{E}\left(|X|^3+3|X|^2|X'|+3|X||X'|^2+|X'|^3\right)\\ &= 2\left(\mathbb{E}\left(|X|^3\right)+3\mathbb{E}\left(|X|\right)\mathbb{E}\left(|X|^2\right)\right) \leqslant 2\left(\mathbb{E}\left(|X|^3\right)+3\sigma^3\right)\end{aligned}$$

so that

$$\mathbb{E}\left(|\Delta W|^3\right) = \mathbb{E}\left(\left|X_I - X_I'\right|^3\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(\left|X_k - X_k'\right|^3\right) \leqslant \frac{2}{n}\sum_{k=1}^{n}\left(\mathbb{E}\left(|X_k|^3\right)+3\sigma_k^3\right)$$

Moreover,

$$\mathbb{E}\left((X-X')^2\middle|X\right) = X^2+\sigma^2$$

thus, with the exchangeable pair of the example 4.3.2 (with $a = 1/n$)

$$\begin{aligned}\left\|1-\frac{1}{2a}\mathbb{E}\left((\Delta W)^2\middle|W\right)\right\|_1 &\leqslant \left\|1-\frac{1}{2a}\mathbb{E}\left((\Delta W)^2\middle|W\right)\right\|_2 = \sqrt{\mathbb{E}\left(\left[1-\frac{1}{2a}\mathbb{E}\left((\Delta W)^2\middle|W\right)\right]^2\right)}\\ &\leqslant \sqrt{\mathbb{E}\left(\left[1-\frac{n}{2}\mathbb{E}\left((\Delta W)^2\middle|(X_k)_{k\leqslant n}\right)\right]^2\right)}\\ &= \sqrt{\mathbb{E}\left(\left[1-\frac{n}{2}\mathbb{E}\left(\left(X_I-X_I'\right)^2\middle|(X_k)_{k\leqslant n}\right)\right]^2\right)}\\ &= \sqrt{\mathbb{E}\left(\left[\frac{1}{2}\sum_{k=1}^{n}(X_k^2-\sigma_k^2)\right]^2\right)} = \frac{1}{2}\sqrt{\sum_{k=1}^{n}\left(\mathbb{E}\left(X_k^4\right)-\sigma_k^4\right)}\end{aligned}$$

□

We can look at the special case of sum of i.i.d. random variables to see the accuracy of this bound : we set

$$W_n := \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k$$

with $(Y_k)_k \sim iid(X)$ and $\mathbb{E}(Y_k) = 0$, $\mathbb{E}\left(Y_k^2\right) = 1$. We see that if $X_k := Y_k/\sqrt{n}$ we are in the case of application of the last corollary, getting

$$|\mathbb{E}(h(W)) - \mathbb{E}(h(Z))| \leqslant \|h-\mathbb{E}(h(Z))\|_\infty\sqrt{\frac{\pi}{2}\frac{\mathbb{E}\left(Y^4\right)-1}{n}} + \left\|h'\right\|_\infty\frac{\mathbb{E}\left(|Y|^3\right)+3}{\sqrt{n}}$$

For other applications than independent sequences, the reader can consult [85] or [94] for instance.

4.4 Stein's method with bias

The bias interpretation is an efficient way to compare $\mathbb{E}\left(Wf(W)\right)$ and $\mathbb{E}\left(f'(W)\right)$ in the spirit of the following result : if W is a real random variable with Lebesgue density $f_W > 0$ a.e. and such that $\mathbb{E}\left(X\right) = 0$ and $\mathbb{E}\left(X^2\right) < \infty$, and defining

$$h_W(x) := \frac{\mathbb{E}\left(W\mathbb{1}_{\{W \geqslant x\}}\right)}{f_W(x)}$$

then for φ absolutely continuous such that $\mathbb{E}\left(|W\varphi(W)|\right) < \infty$, we have

$$\mathbb{E}\left(W\varphi(W)\right) = \mathbb{E}\left(h_W(W)\varphi'(W)\right)$$

It is easy to prove that h_W is positive and in $L^1(\mathbb{P}_W)$, so we can define the probability distribution

$$\mathbb{P}_{W^{(0)}} := \frac{h_W(W)}{\mathbb{E}_{\mathbb{P}}\left(h_W(W)\right)} \bullet \mathbb{P}_W$$

This amounts to write

$$\mathbb{E}\left(W\varphi(W)\right) = \mathbb{E}\left(h_W(W)\right)\mathbb{E}\left(\varphi'\left(W^{(0)}\right)\right)$$

Last, taking $\varphi = id$, we have $\mathbb{E}\left(h_W(W)\right) = \mathbb{E}\left(W^2\right) =: \sigma_W^2$ so that,

$$\mathbb{E}\left(W\varphi(W)\right) = \sigma_W^2\mathbb{E}\left(\varphi'\left(W^{(0)}\right)\right)$$

Such a random variable $W^{(0)}$ is said to be the *zero-bias transform* of W. This is a particular case of change of probability in the same vein as the size-bias transform defined in (3.14).

4.4.1 The size-bias coupling

We recall the following

Definition 4.4.1. Let $X \geqslant 0$ be a random variable with expectation $\mu := \mathbb{E}\left(X\right) < \infty$. A random variable X^s is said to be a *size-bias transform* of X if, for all real functions f such that $\mathbb{E}\left(|Xf(X)|\right) < \infty$, we have

$$\mathbb{E}\left(Xf(X)\right) = \mu\mathbb{E}\left(f\left(X^s\right)\right)$$

An equivalent definition is thus

$$\mathbb{P}_{X^s} := \frac{X}{\mathbb{E}\left(X\right)} \bullet \mathbb{P}_X \tag{4.14}$$

In the same vein as in the case of an exchangeable pair, we also have an abstract approximation theorem using Stein's method, but with the couple (X, X^s) in place of an exchangeable pair. The philosophy is exactly the same : the difference $\Delta X := X^s - X$ should be thought of as a *small perturbation* of X.

Theorem 4.4.2 (Stein's bound with size-bias coupling). *Let $X \geqslant 0$ be a random variable with expectation $\mu := \mathbb{E}(X) < \infty$ and* $\mathrm{Var}X =: \sigma^2 < \infty$. *Let X^s be a size-bias transform of X defined on the same probability space as X. Then, if $W := (X-\mu)/\sigma$, $Z \sim \mathscr{N}(0,1)$, and setting $\Delta X := X^s - X$, we have, for all f continuous such that $\|f'\|_\infty < \infty$ and $\|f''\|_\infty < \infty$*

$$|\mathbb{E}(\mathcal{A}f(W))| \leqslant \|f'\|_\infty \left\|1 - \frac{\mu}{\sigma^2}\mathbb{E}(\Delta X|X)\right\|_1 + \|f''\|_\infty \frac{\mu}{\sigma^3}\|\Delta X\|_2^2 \tag{4.15}$$

Proof. We have the following *polarisation identity* which explains why it is interesting to consider such a couple

$$\mathbb{E}(Wf(W)) = \mathbb{E}\left(\frac{X-\mu}{\sigma} f\left(\frac{X-\mu}{\sigma}\right)\right) = \frac{\mu}{\sigma}\mathbb{E}\left[f\left(\frac{X^s-\mu}{\sigma}\right) - f\left(\frac{X-\mu}{\sigma}\right)\right] \tag{4.16}$$

Now, we can write a Taylor development for the function $f_{\mu,\sigma} : x \mapsto f((X-\mu)/\sigma)$

$$\begin{aligned}
\mathbb{E}(Wf(W)) &= \frac{\mu}{\sigma}\mathbb{E}[f_{\mu,\sigma}(X^s) - f_{\mu,\sigma}(X)] = \frac{\mu}{\sigma}\mathbb{E}[f_{\mu,\sigma}(X+\Delta X) - f_{\mu,\sigma}(X)] \\
&= \frac{\mu}{\sigma}\mathbb{E}\left[\Delta X f'_{\mu,\sigma}(X) + \int_X^{X+\Delta X}(t-X)f''_{\mu,\sigma}(t)dt\right] \\
&= \frac{\mu}{\sigma}\mathbb{E}\left[\frac{\Delta X}{\sigma}f'(W) + \int_X^{X+\Delta X}\frac{(t-X)}{\sigma^2}f''\left(\frac{t-\mu}{\sigma}\right)dt\right]
\end{aligned}$$

Hence,

$$\begin{aligned}
|\mathbb{E}(\mathcal{A}f(W))| &= |\mathbb{E}(f'(W) - Wf(W))| \\
&= \left|\mathbb{E}\left(f'(W) - f'(W)\frac{\mu}{\sigma^2}\Delta X + f'(W)\frac{\mu}{\sigma^2}\Delta X - Wf(W)\right)\right| \\
&= \left|\mathbb{E}\left(f'(W)\left[1 - \frac{\mu}{\sigma^2}\Delta X\right]\right) - \frac{\mu}{\sigma}\int_X^{X+\Delta X}\frac{(t-X)}{\sigma^2}f''\left(\frac{t-\mu}{\sigma}\right)dt\right| \\
&\leqslant \|f'\|_\infty \mathbb{E}\left(\left|1 - \frac{\mu}{\sigma^2}\mathbb{E}(\Delta X|X)\right|\right) + \|f''\|_\infty \frac{\mu}{\sigma^3}\mathbb{E}\left(\left|\int_X^{X+\Delta X}|t-X|\,dt\right|\right) \\
&= \|f'\|_\infty \left\|1 - \frac{\mu}{\sigma^2}\mathbb{E}(\Delta X|X)\right\|_1 + \|f''\|_\infty \frac{\mu}{\sigma^3}\frac{1}{2}\mathbb{E}((\Delta X)^2)
\end{aligned}$$

$\square$

Of course, using Stein's estimates (4.44), we can deduce

$$|\mathbb{E}(h(W)) - \mathbb{E}(h(Z))| \leqslant \|h_\Phi\|_\infty \sqrt{\frac{\pi}{2}}\frac{\mu}{\sigma^2}\sqrt{\mathrm{Var}\mathbb{E}(\Delta X|X)} + \|h'\|_\infty \frac{\mu}{\sigma^3}\mathbb{E}\left((\Delta X)^2\right)$$

and its direct corollary

$$W_1(W,Z) \leqslant \sqrt{\frac{\pi}{2}}\frac{\mu}{\sigma^2}\sqrt{\mathrm{Var}\mathbb{E}(\Delta X|X)} + \frac{\mu}{\sigma^3}\mathbb{E}\left((\Delta X)^2\right)$$

Here, we have used the inequality

$$\left\|1-\frac{\mu}{\sigma^2}\mathbb{E}\left(\Delta X|X\right)\right\|_1 \leqslant \left\|1-\frac{\mu}{\sigma^2}\mathbb{E}\left(\Delta X|X\right)\right\|_2$$
$$=\frac{\mu}{\sigma^2}\sqrt{\mathbb{E}\left(\left[\mathbb{E}\left(\Delta X|X\right)-\frac{\sigma^2}{\mu}\right]^2\right)}=\frac{\mu}{\sigma^2}\sqrt{\mathrm{Var}\mathbb{E}\left(\Delta X|X\right)}$$

and the fact that

$$\mathbb{E}\left(X^s\right)=\frac{\mathbb{E}\left(X^2\right)}{\mathbb{E}\left(X\right)}=\frac{\sigma^2+\mu^2}{\mu}$$

Like for the exchangeable pair philosophy, it is accurate to ask how to couple a random variable X to one of its size-bias transformation X^s. For this, we take the case of the "canonical" example, namely, the sum of random variables, that we do not suppose independent this time.

Example 4.4.3. Let $X := \sum_{k=1}^n X_k$ with $X_k \geqslant 0$ and $\mathbb{E}\left(X_k\right) = \mu_k$. The following receipe allows to construct a size-bias of X :

(i) For all $i \in [\![1,n]\!]$, let X_i^s be the size-bias of X_i independent of $(X_j)_{j\neq i}$ and $(X_j^s)_{j\neq i}$ and define $\left(X_j^{(i)}\right)_{j\neq i}$ such that

$$\left(\left(X_j^{(i)}\right)_{j\neq i}\middle|X_i^s\right) \stackrel{\mathcal{L}}{=} \left(\left(X_j\right)_{j\neq i}\middle|X_i\right)$$

(ii) With $\mu := \mathbb{E}\left(X\right) = \mathbb{E}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \mu_k$, set

$$I \sim \sum_{k=1}^n \frac{\mu_k}{\mu}\delta_k$$

that is : $\mathbb{P}\left(I=k\right) = \mu_k/\mu$ and $I \in [\![1,n]\!]$.

(iii) Let I with such a distribution being independent of all the random variables defined. Define

$$X^s := \sum_{j\neq I} X_j^{(I)} + X_I^s$$

Proposition 4.4.4. *Let $X := \sum_{k=1}^n X_k$ with $X_k \geqslant 0$ and $\mathbb{E}\left(X_k\right) = \mu_k$, $\mathbb{E}\left(X\right) =: \mu = \sum_{k=1}^n \mu_k$. Then, X^s constructed as above is equal in law to a size-bias distribution of X.*

Proof. Let $\boldsymbol{X} := (X_1, \ldots, X_n)$ and $\boldsymbol{X}^i$ the vector written in the canonical basis $(e_1, \ldots, e_n)$

$$\boldsymbol{X}^i = \boldsymbol{X} + \left(X_j^{(i)} - X_j\right) e_j + (X_i^s - X_i)\, e_i$$

for $i \neq j$. To prove the result, it is enough to show

$$\mathbb{E}\left(X f(\boldsymbol{X})\right) = \mu \mathbb{E}\left(f\left(\boldsymbol{X}^I\right)\right)$$

for all $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that $\mathbb{E}\left(|X f(\boldsymbol{X})|\right) < \infty$. As $\mu \mathbb{E}\left(f\left(\boldsymbol{X}^I\right)\right) = \sum_{i=1}^n \mu_i \mathbb{E}\left(f\left(\boldsymbol{X}^i\right)\right)$, to prove this last result, it is enough to show that for all $i \in [\![1, n]\!]$

$$\mathbb{E}\left(X_i f(\boldsymbol{X})\right) = \mu_i \mathbb{E}\left(f\left(\boldsymbol{X}^i\right)\right) \tag{4.17}$$

Summing on $i \in [\![1, n]\!]$ will then give the last equation. To prove (4.17), set $h(X_i) := \mathbb{E}\left(f(\boldsymbol{X}) | X_i\right)$. We have

$$\mathbb{E}\left(X_i f(\boldsymbol{X})\right) = \mathbb{E}\left(X_i h(X_i)\right) = \mu_i \mathbb{E}\left(h(X_i^s)\right) = \mu_i \mathbb{E}\left(\mathbb{E}\left(f(\boldsymbol{X}^i) | X_i\right)\right) = \mu_i \mathbb{E}\left(f\left(\boldsymbol{X}^i\right)\right)$$

Q.E.D.

□

As a corollary, we note the case of independent $(X_k)_k$: if X_i^s has the size-bias distribution of X_i and is independent of $(X_j)_{j \neq i}$ and $(X_j^s)_{j \neq i}$, then, with I previously defined, $X^s := X - X_I + X_I^s$ has the law of a size-bias distribution of X.

Application : Using the previous construction of a size-bias couple and the estimate (4.15), we can bound the L^1-Wasserstein distance between a normalised sum of independent random variables and a Gaussian. This is left as an exercise.

4.4.2 The zero-bias coupling

Same philosophy as before, but with X that is not supposed positive and such that $\mathbb{E}(X) = 0$, $\mathbb{E}\left(X^2\right) < \infty$.

Definition 4.4.5. Let $X \geqslant 0$ be a random variable with expectation $\mathbb{E}(X) = 0$ and variance $\sigma^2 := \mathbb{E}\left(X^2\right) < \infty$. A random variable $X^{(0)}$ is said to be a *zero-bias transform* of X if, for all real functions f such that $\mathbb{E}\left(|X f(X)|\right) < \infty$, we have

$$\mathbb{E}\left(X f(X)\right) = \sigma^2 \mathbb{E}\left(f'\left(X^{(0)}\right)\right) \tag{4.18}$$

The zero-bias transform is a distributional transform, hence, it that can be written as a change of measure with a certain density, but not necessarily with respect to the initial measure, as one can see on this example :

Example 4.4.6. Let $B \in \{\pm 1\}$ be a symmetric Bernoulli random variable of parameter $1/2$. The zero-bias transform $B^{(0)}$ is $\mathscr{U}([-1,1])$-distributed, as shown by the following computation

$$i\theta \mathbb{E}\left(e^{i\theta B^{(0)}}\right) = \mathbb{E}\left(Be^{i\theta B}\right) = \frac{1}{2}\left(-e^{-i\theta} + e^{i\theta}\right)$$

which implies

$$\mathbb{E}\left(e^{i\theta B^{(0)}}\right) = \frac{1}{2i\theta}\left(e^{i\theta} - e^{-i\theta}\right) = \int_{-1}^{1} e^{i\theta u}\frac{du}{2} = \mathbb{E}\left(e^{i\theta U_{[-1,1]}}\right)$$

Hence, $\mathbb{P}_B = \frac{1}{2}(\delta_{-1} + \delta_1)$ but $\mathbb{P}_{B^{(0)}}(dx) = \frac{1}{2}\mathbb{1}_{\{|x|\leqslant 1\}}dx$ and we change from a singular to an absolutely continuous measure (with respect to the Lebesgue measure).

Note that the passage from $\{-1,1\}$ to $[-1,1]$ is an interpolation, and that the zero-bias transform consists in a smoothing of the distribution.

Remark 4.4.7. The zero bias transform has as a unique fixed point the normal distribution : this is a reformulation of the Stein's equation in the case of $\mathscr{N}(0,\sigma^2)$, with a Stein operator $\mathcal{A}_\sigma f(x) = f'(x) - \sigma^2 x f(x)$. Thus, we can see that Stein's method can be understood as a **fixed point problem** in distribution, wich is a classical idea in analysis.

It is not clear wether the zero-bias transform defines a probabiliy distribution. We will see later that this is indeed the case.

One useful property of such a transform is the following identity in distribution (c.f. [44, 85]) : if $S_n = \sum_{k=1}^n X_k$ with $(X_k)_k$ a sequence of i.i.d. random variables, then

$$S_n^{(0)} \overset{\mathcal{L}}{=} S_n + \left(X_I^{(0)} - X_I\right) = \sum_{k\neq I} X_k + X_I^{(0)} \tag{4.19}$$

where $I \sim \mathscr{U}(\llbracket 1,n\rrbracket)$ is a random variable independent of $(X_k)_k$ and $(X_k^{(0)})_k$, those two last sequences being independent, and $(X_k^{(0)})_k$ being a sequence of i.i.d. random variables distributed according to the zero-bias distribution of X.

The proof is straightforward and is a replica of the proof of lemma 3.4.5 : for f a differentiable function with bounded derivatives and $S_n^{\langle -k\rangle} := \sum_{\ell\neq k} X_\ell$ we have, by independence,

$$\begin{aligned}
\mathbb{E}\left(f'\left(S_n^{(0)}\right)\right) &:= \frac{1}{\mathbb{E}\left(S_n^2\right)}\mathbb{E}\left(S_n f(S_n)\right) = \frac{1}{n}\sum_{k=1}^n \mathbb{E}\left(X_k f(S_n)\right)\\
&= \frac{1}{\mathbb{E}\left(S_n^2\right)}\sum_{k=1}^n \mathbb{E}\left(X_k f\left(S_n^{\langle -k\rangle} + X_k\right)\right)\\
&= \frac{1}{\mathbb{E}\left(S_n^2\right)}\sum_{k=1}^n \mathbb{E}\left(X_k^2\right)\mathbb{E}\left(f'\left(S_n^{\langle -k\rangle} + Y_k^{(0)}\right)\right)\\
&= \mathbb{E}\left(f'\left(S_n^{\langle -I\rangle} + Y_I^{(0)}\right)\right) = \mathbb{E}\left(f'\left(S_n - X_I + Y_I^{(0)}\right)\right)
\end{aligned}$$

where we have set $\mathbb{P}(I = k) = \mathbb{E}\left(X_k^2\right)/\mathbb{E}\left(S_n^2\right)$, which gives the uniform distribution in the case of a sequence of i.i.d. random variables.

Like in the case of the size-bias transform and the exchangeable pair, we have an approximation theorem :

Theorem 4.4.8 (Stein's bound with zero-bias coupling)**.** *Let W such that $\mathbb{E}(W) = 0$ and $\mathbb{E}(W^2) = 1$. Let $W^{(0)}$ be a zero-bias transform of W defined on the same probability space as W. Then, if $Z \sim \mathcal{N}(0,1)$, and setting $\Delta W := W^{(0)} - W$, we have, for all f continuous such that $\|f''\|_\infty < \infty$*

$$|\mathbb{E}(\mathcal{A}f(W))| \leqslant \|f''\|_\infty \|\mathbb{E}(\Delta W|W)\|_1 \tag{4.20}$$

Proof. We have

$$|\mathbb{E}(\mathcal{A}f(W))| = \left|\mathbb{E}\left(f'(W) - Wf(W)\right)\right| = \left|\mathbb{E}\left(f'(W) - f'\left(W^{(0)}\right)\right)\right| \leqslant \|f''\|_\infty \mathbb{E}\left(\left|W - W^{(0)}\right|\right)$$

□

As a corollary and using the Stein's estimates, we have

$$W_1(W,Z) \leqslant 2\mathbb{E}\left(\left|W - W^{(0)}\right|\right)$$

This bound is extremely simple, but the price to pay is the construction of such a coupling. For independent random variables, the latest construction applies with a few changes (see [85] for instance).

To conclude, we give some properties of the zero-bias transform that show it is well defined.

Proposition 4.4.9. *Let W such that $\mathbb{E}(W) = 0$ and $\mathbb{E}(W^2) = 1$.*

1. *There is a unique probability distribution for a random variable $W^{(0)}$ such that for all f absolutely continuous satisfying $\mathbb{E}(|Wf(W)|) < \infty$*

$$\mathbb{E}(Wf(W)) = \mathbb{E}\left(f'\left(W^{(0)}\right)\right)$$

2. *The distribution of $W^{(0)}$ is absolutely continuous with respect to Lebesgue measure and has density*

$$f_{W^{(0)}}(x) = \mathbb{E}\left(W\mathbb{1}_{\{W \geqslant x\}}\right)$$

Proof. We have, by Fubini and for f such that $f(0) = 0$

$$\begin{aligned}\int_0^{+\infty} f'(u)\mathbb{E}\left(W\mathbb{1}_{\{W\geqslant u\}}\right)du &= \mathbb{E}\left(W\int_0^{+\infty} f'(u)\mathbb{1}_{\{W\geqslant u\}}du\right) = \mathbb{E}\left(W\int_0^W f'(u)du\mathbb{1}_{\{W\geqslant 0\}}\right)\\ &= \mathbb{E}\left(f(W)W\mathbb{1}_{\{W\geqslant 0\}}\right)\end{aligned}$$

and similarly $\int_{-\infty}^{0} f'(u)\mathbb{E}\left(W\mathbb{1}_{\{W\geqslant u\}}\right)du = \mathbb{E}\left(f(W)W\mathbb{1}_{\{W\leqslant 0\}}\right)$. Thus, summing, we have

$$\int_{-\infty}^{+\infty} f'(u)\mathbb{E}\left(W\mathbb{1}_{\{W\geqslant u\}}\right)du = \mathbb{E}\left(f(W)W\right) \tag{4.21}$$

The validity of such a formula is for functions f such that $\mathbb{E}\left(|f(W)W|\right) < \infty$ that are absolutely continuous. The last formula defines a function $f_{W^{(0)}}$. Let us prove that this is a probability density.

$f_{W^{(0)}}$ is positive as for $x > 0$, $\mathbb{E}\left(W\mathbb{1}_{\{W\geqslant x\}}\right) \geqslant x\mathbb{P}(W \geqslant x) > 0$. And for $x < 0$, that is $-x > 0$,

$$\begin{aligned}\mathbb{E}\left(W\mathbb{1}_{\{W\geqslant x\}}\right) &= \mathbb{E}\left(W\left[1-\mathbb{1}_{\{W\leqslant x\}}\right]\right) = \mathbb{E}\left(W\right) - \mathbb{E}\left(W\mathbb{1}_{\{W\leqslant x\}}\right)\\ &= \mathbb{E}\left(-W\mathbb{1}_{\{-W\geqslant -x\}}\right) \quad \text{as } \mathbb{E}(W) = 0\\ &\geqslant -x\mathbb{P}\left(-W\geqslant -x\right) > 0\end{aligned}$$

Last,

$$\begin{aligned}\int_{0}^{+\infty} f_{W^{(0)}}(x)dx &= \int_{0}^{+\infty}\mathbb{E}\left(W\mathbb{1}_{\{W\geqslant x\}}\right)dx = \mathbb{E}\left(W\int_{0}^{+\infty}\mathbb{1}_{\{W\geqslant x\}}dx\right)\\ &= \mathbb{E}\left(W\int_{0}^{W}dx\mathbb{1}_{\{W\geqslant 0\}}\right) = \mathbb{E}\left(W^2\mathbb{1}_{\{W\geqslant 0\}}\right)\end{aligned}$$

and similarly

$$\int_{-\infty}^{0} f_{W^{(0)}}(x)dx = \mathbb{E}\left(W^2\mathbb{1}_{\{W\leqslant 0\}}\right)$$

Summing and using $\mathbb{E}\left(W^2\right) = 1$ gives the desired result. Let us remark that taking $f(W) = W$ in (4.21) gives directly the result.

□

4.5 Stein's method : a summary

The conventions on the functions f are those given in the theorems. We set ΔW for the perturbation $W - W^*$, with $W^* \in \left\{W', W^s, W^{(0)}\right\}$ and $\Delta f(W) := f(W^*) - f(W) = f(W + \Delta W) - f(W)$. The results that are obtained by means of the abstract approximation theorems are summarized in the following table

Couples	**Polarisation**	**Estimates**
Exchangeable pair (W, W')	$\mathbb{E}(Wf(W)) = \frac{1}{2a}\mathbb{E}(\Delta W \Delta f(W))$	$\lvert\mathbb{E}(\mathcal{A}f(W))\rvert \leqslant \lVert f'\rVert_\infty \left\lVert 1 - \frac{1}{2a}\mathbb{E}\left((\Delta W)^2 \middle\vert W\right)\right\rVert_1 + \lVert f''\rVert_\infty \frac{1}{4a}\lVert\Delta W\rVert_3^3$
Size-bias pair (W, W^s)	$\mathbb{E}(Wf(W)) = \frac{\mu}{\sigma}\mathbb{E}\left(\Delta f\left(\frac{W-\mu}{\sigma}\right)\right)$	$\lvert\mathbb{E}(\mathcal{A}f(W))\rvert \leqslant \lVert f'\rVert_\infty \left\lVert 1 - \frac{\mu}{\sigma}\mathbb{E}(\Delta W \vert W)\right\rVert_1 + \lVert f''\rVert_\infty \frac{\mu}{\sigma}\lVert\Delta W\rVert_2^2$
Zero-bias pair $(W, W^{(0)})$	$\mathbb{E}(Wf(W)) = \mathbb{E}\left(f'(W^{(0)})\right)$	$\lvert\mathbb{E}(\mathcal{A}f(W))\rvert \leqslant \lVert f''\rVert_\infty \lVert\mathbb{E}(\Delta W \vert W)\rVert_1$

Moreover, as noticed in remark 4.4.7, if Stein's method transfers the information on the distribution to approximate to its characteristic operator, escaping then from the domain of probability theory, it comes back to it by re-translating Stein's equation into a fixed point equation in law whose only attractor is given by the law to approximate.

To formalise what is Stein's method, consider an abstract operator $\mathcal{A}$ such that

$$\mathcal{A} = \mathcal{L}_1 - \mathcal{L}_2$$

where $\mathcal{L}_2$ must be understood as a perturbation of $\mathcal{L}_1$, i.e. $\mathcal{L}_2 = \Delta\mathcal{A}$ and $\mathcal{L}_1 = \mathcal{A} + \Delta\mathcal{A}$.

Supposing that $\mathcal{A}$ is the characteristic operator of a distribution according to which the random variable X is distributed, i.e. $\mathbb{E}(\mathcal{A}f(X)) = 0$ for f in a good class of test functions, and supposing that one can define X^* via *the polarisation identity*

$$\mathbb{E}(\mathcal{L}_2 f(X)) = \mathbb{E}(\mathcal{L}_1 f(X^*))$$

then,

$$\mathbb{E}(\mathcal{A}f(X)) = \mathbb{E}(\mathcal{L}_1 f(X) - \mathcal{L}_2 f(X)) = \mathbb{E}(\mathcal{L}_1 f(X) - \mathcal{L}_1 f(X^*)) \leqslant \lVert D\mathcal{L}_1 f\rVert_\infty \mathbb{E}(\lvert X - X^*\rvert)$$

In Stein's method, one writes

$$\mathbb{E}(h(X)) - \mathbb{E}(h(Z)) = \mathbb{E}(h(X) - \mathbb{E}(h(Z))) = \mathbb{E}(h_\Phi(X)) = \mathbb{E}\left(\mathcal{A}\mathcal{A}^{-1}h_\Phi(X)\right)$$

where $\mathcal{A}^{-1}$ is the pseudo-inverse of $\mathcal{A}$ and where

$$h_\Phi(x) := h(x) - \mathbb{E}(h(Z))$$

One can thus write

$$\lvert\mathbb{E}(h(X)) - \mathbb{E}(h(Z))\rvert = \left\lvert\mathbb{E}\left(\mathcal{A}\mathcal{A}^{-1}h_\Phi(X)\right)\right\rvert \leqslant \left\lVert D\mathcal{L}_1\mathcal{A}^{-1}h_\Phi\right\rVert_\infty \mathbb{E}(\lvert X - X^*\rvert)$$

Since $\mathcal{L}_1 = \Delta\mathcal{A}$ is the perturbation, it is *transfered* to X^* by the probabilistic transformation. If $\Delta\mathcal{A}$ is a small perturbation, one can expect that $\Delta X := X - X^*$ is again a small perturbation. One first needs to bound

$$\left\|D\mathcal{L}_1\mathcal{A}^{-1}h_\Phi\right\|_\infty = \left\|D(\mathcal{A}+\Delta\mathcal{A})\mathcal{A}^{-1}h_\Phi\right\|_\infty = \left\|D(\mathcal{A}\mathcal{A}^{-1}+\Delta\mathcal{A}\mathcal{A}^{-1})h_\Phi\right\|_\infty = \left\|h' + \Delta\mathcal{A}\mathcal{A}^{-1}h_\Phi\right\|_\infty$$

which can be done in a standard way once the explicit expression of $\mathcal{A}$ is specified. One needs then to bound $|X - X^*|$ in L^1 (or L^p). These two random variables not being a priori in the same probability space, a suitable coupling between them must be found ; this is where the specificity of the problem arises.

Note that an abstract random variable equal in distribution to X^* can be defined in the case where X has a Lebesgue-density $f_X \in L^2(\lambda)$ (with λ the Lebesgue measure) :

$$\mathbb{E}\left(\mathcal{L}_2 h(X)\right) = \int_{\mathbb{R}} \mathcal{L}_2 h\ f_X d\lambda = \langle \mathcal{L}_2 h, f_X\rangle_{L^2} = \langle \mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_1 h, f_X\rangle_{L^2} = \langle \mathcal{L}_1 h, \left(\mathcal{L}_2\mathcal{L}_1^{-1}\right)^* f_X\rangle_{L^2}$$

Thus, supposing

$$h_X := \frac{\left(\mathcal{L}_2\mathcal{L}_1^{-1}\right)^* f_X}{f_X} \geqslant 0$$

we get

$$\mathbb{E}\left(\mathcal{L}_2 h(X)\right) = \mathbb{E}\left(h_X(X)\mathcal{L}_1 h(X)\right)$$

and we can set

$$\mathbb{P}_{X^*} := \frac{h_X(X)}{\mathbb{E}\left(h_X(X)\right)} \bullet \mathbb{P}_X$$

Finally, Stein's method can be seen as the conjunction of three principles :

- a **fixed point principle** that characterises the law to approximate by means of a probabilistic transformation equivalent to computing the kernel of a certain operator,
- a **perturbation principle** that consists in perturbating the characteristic operator which is equivalent to perturbate the random variable with a control of the perturbation in order to stay in the domain of attraction of the law to approximate,
- a **transfer principle** that consists in transfering the characterisation of the law and the perturbation of the operator from the world of operators to the domain of probability theory, giving an equation in law and a probabilistic perturbation.

Note that, somehow, methods involving transforms such as the Fourier-Laplace or the Stieltjes ones consist in passing from random variables to functions, and aim at proving a fixed point equation : the Gaussian distribution is the fixed point of the Fourier transform, since the Gaussian density is equal to its Fourier transform, and a fixed point equation is the key to the proof of the semi-circle distribution using the Stieltjes transform. The sucess of fixed point methods and perturbative methods may thus explain the sucess of Stein's method and its wide variety of application.

4.6 Stein's method and mod-Gaussian convergence

As seen in theorem 3.4.7, mod-* convergence can be interpreted, under some restrictions, as a second-order distributional approximation by means of a certain random variable involving the limiting mod-Laplace function Φ. It hence becomes relevant to find a bound in a smooth metric or in the Kolmogorov metric to the approximation of the sequence converging in the mod-* sense and its mod-* approximation, and Stein's methodology can be of some use.

In what follows, we restrict ourselves to the Gaussian case, being understood that the Poisson case can be treated in the same way.

4.6.1 A Stein's operator for the penalised Gaussian distribution

In order to apply Stein's method to a sequence of random variables $(W_n)_n$ converging in the mod-Gaussian sense with parameters $((\gamma_n)_n, \Phi)$, we want a characteristic operator $\mathcal{L}_{H_\gamma}$ of $H_\gamma \sim \mathscr{H}(\Phi, \gamma)$.

Theorem 4.6.1 (Characteristic operator for a penalisation of the Gaussian distribution)**.** *Let Φ satisfying the hypotheses of theorem 3.4.7 and $H_\gamma \sim \mathscr{H}(\Phi, \gamma)$. Suppose that $\Phi > 0$ on $\mathbb{R}$ and set*

$$\begin{aligned}
\Psi(x) &:= \log \Phi(x) \\
\kappa_\gamma(x) &:= \frac{x^2}{2\gamma^2} - \Psi\left(\frac{x}{\gamma^2}\right) \\
\rho_\gamma(x) &:= \frac{x}{\gamma^2} - \frac{1}{\gamma^2}\Psi'\left(\frac{x}{\gamma^2}\right) = \kappa'_\gamma(x) \\
\widetilde{\mathcal{H}}_{\Phi,\gamma} &:= \left\{h \in \mathcal{C}^1_m \ / \ \mathbb{E}\left(\left|h'(H_\gamma)\right|\right) < \infty, \lim{}_{\pm\infty} h = 0\right\}
\end{aligned}$$

where $\mathcal{C}^1_m$ is the space of continuous and piecewise continuously differentiable functions $f : \mathbb{R} \to \mathbb{R}$.

Suppose moreover that

1. $\kappa_\gamma(x) \xrightarrow[x \to \pm\infty]{} +\infty$,
2. $\forall\, x \in \mathbb{R}$, $\rho'_\gamma(x) \geqslant 0$.

Then, a characteristic operator of H_γ on $\widetilde{\mathcal{H}}_{\Phi,\gamma}$ is $\mathcal{L}_{\Phi,\gamma}$ defined for all $h \in \widetilde{\mathcal{H}}_{\Phi,\gamma}$ by

$$\mathcal{L}_{\Phi,\gamma} h(x) := h'(x) - \left(\frac{x}{\gamma^2} - \frac{1}{\gamma^2}\Psi'\left(\frac{x}{\gamma^2}\right)\right) h(x) \tag{4.22}$$

Proof. By inverting the first order differential operator $\mathcal{L}_{\Phi,\gamma} = D - \rho_\gamma$, it is easily proven that for functions $h_\gamma := h - \mathbb{E}\left(h(H_\gamma)\right)$ with $h \in \widetilde{\mathcal{H}}_{\Phi,\gamma}$, we have

$$\mathcal{L}_{\Phi,\gamma} g = h_\gamma \iff g(x) = \mathcal{L}^{-1}_{\Phi,\gamma} h_\gamma(x) = \frac{1}{f_{H_\gamma}(x)} \int_{-\infty}^{x} f_{H_\gamma}(y) h_\gamma(y) dy$$

f_{H_γ} being given in (3.20), i.e. $f_{H_\gamma} = e^{-\kappa_\gamma}/c_\gamma$. Using the random variable H_γ, we can write

$$\mathcal{L}_{\Phi,\gamma}^{-1} h_\gamma(x) = \frac{\mathbb{E}\left(h_\gamma(H_\gamma)\mathbb{1}_{\{H_\gamma \leqslant x\}}\right)}{f_{H_\gamma}(x)} = -\frac{\mathbb{E}\left(h_\gamma(H_\gamma)\mathbb{1}_{\{H_\gamma \geqslant x\}}\right)}{f_{H_\gamma}(x)} \tag{4.23}$$

the last equality coming from the fact that $\mathbb{E}\left(h_\gamma(H_\gamma)\right) = 0$.

Now, if a random variable Y is such that for all $h \in \widetilde{\mathcal{H}}_{\Phi,\gamma}$, $\mathbb{E}\left(\mathcal{L}_{\Phi,\gamma} h(Y)\right) = 0$, it is true in particular for the function

$$h_{x,\gamma} := \mathcal{L}_{\Phi,\gamma}^{-1}(u_x - \mathbb{E}\left(u_x(H_\gamma)\right)) \quad \text{with} \quad u_x : y \mapsto \mathbb{1}_{\{y \leqslant x\}}$$

Let us prove that $h_{x,\gamma} \in \widetilde{\mathcal{H}}_{\Phi,\gamma}$. By (4.23), we can write

$$h_{x,\gamma}(y) = \frac{\mathbb{E}\left(\left(\mathbb{1}_{\{H_\gamma \leqslant x\}} - \mathbb{P}\left(H_\gamma \leqslant x\right)\right)\mathbb{1}_{\{H_\gamma \leqslant y\}}\right)}{f_{H_\gamma}(y)} = \frac{\operatorname{cov}\left(\mathbb{1}_{\{H_\gamma \leqslant x\}}, \mathbb{1}_{\{H_\gamma \leqslant y\}}\right)}{f_{H_\gamma}(y)}$$

Let $x \geqslant 0$. As $H_\gamma \stackrel{\mathcal{L}}{=} -H_\gamma$, we only have to consider this case.

$$\begin{aligned}
\mathbb{P}\left(H_\gamma \geqslant x\right) &= \int_x^{+\infty} e^{-\kappa_\gamma(u)} \frac{du}{c_\gamma} \quad \text{with } c_\gamma \text{ given in (3.20)} \\
&\leqslant \int_x^{+\infty} \frac{\kappa_\gamma'(u)}{\kappa_\gamma'(x)} e^{-\kappa_\gamma(u)} \frac{du}{c_\gamma} \quad \text{since } \kappa_\gamma''(x) = \rho_\gamma'(x) \geqslant 0 \text{ by hypothesis (1)} \\
&= \frac{e^{-\kappa_\gamma(x)}}{c_\gamma \kappa_\gamma'(x)}
\end{aligned}$$

We can rewrite this inequality as

$$\mathbb{P}\left(H_\gamma \geqslant x\right) \leqslant \frac{f_{H_\gamma}(x)}{\rho_\gamma(x)} \tag{4.24}$$

Hence, setting $u_{x,\gamma}(y) := u_x(y) - \mathbb{E}\left(u_x(H_\gamma)\right) = \mathbb{1}_{\{y \leqslant x\}} - \mathbb{P}\left(H_\gamma \leqslant x\right) \leqslant 2$, we have

$$|h_{x,\gamma}(y)| = \left|\mathcal{L}_{\Phi,\gamma}^{-1} u_{x,\gamma}(y)\right| \leqslant \frac{\mathbb{E}\left(|u_{x,\gamma}(H_\gamma)|\,\mathbb{1}_{\{H_\gamma \leqslant y\}}\right)}{f_{H_\gamma}(y)} \leqslant 2\frac{\mathbb{P}\left(H_\gamma \leqslant y\right)}{f_{H_\gamma}(y)} \leqslant \frac{2}{\kappa_\gamma'(y)} \xrightarrow[y \to +\infty]{} 0$$

as $\kappa_\gamma(x) \xrightarrow[x \to \pm\infty]{} +\infty$ by hypothesis (2).

Using $\mathcal{L}_{\Phi,\gamma} = D - \rho_\gamma$ and the Stein's equation $\mathcal{L}_{\Phi,\gamma} h_{x,\gamma} = u_{x,\gamma}$, we have

$$Dh_{x,\gamma} = \rho_\gamma h_{x,\gamma} + u_{x,\gamma}$$

which implies

$$|Dh_{x,\gamma}(y)| \leqslant |u_{x,\gamma}(y)| + |\rho_\gamma(y) h_{x,\gamma}(y)| \leqslant 2\left(1 + \left|\rho_\gamma(y)\frac{\mathbb{P}\left(H_\gamma \geqslant y\right)}{f_{H_\gamma}(y)}\right|\right) \leqslant 4 \quad \text{by (4.24)}$$

which proves that $\mathbb{E}\left(\left|h'_{x,\gamma}(H_\gamma)\right|\right) < \infty$, i.e. $h_{x,\gamma} \in \widetilde{\mathcal{H}}_{\Phi,\gamma}$. Then, for all $x \in \mathbb{R}$,

$$0 = \mathbb{E}\left(\mathcal{L}_{\Phi,\gamma} h_{x,\gamma}(Y)\right) = \mathbb{E}\left(u_x(Y) - \mathbb{E}\left(u_x(H_\gamma)\right)\right) = \mathbb{P}\left(Y \leqslant x\right) - \mathbb{P}\left(H_\gamma \leqslant x\right)$$

that is : $Y \overset{\mathcal{L}}{=} H_\gamma$.

Reciprocally, let us prove that for all $h \in \widetilde{\mathcal{H}}_{\Phi,\gamma}$,

$$\mathbb{E}\left(\mathcal{L}_{\Phi,\gamma} h(H_\gamma)\right) = 0$$

As $-\rho_\gamma = D\log(f_{H_\gamma}) = f'_{H_\gamma}/f_{H_\gamma}$ we have, by the Fubini theorem

$$\begin{aligned}
\mathbb{E}\left(h'(H_\gamma)\right) &= \int_{\mathbb{R}} h'(u) f_{H_\gamma}(u)du = \int_{-\infty}^{0} h'(u) f_{H_\gamma}(u)du + \int_{0}^{+\infty} h'(u) f_{H_\gamma}(u)du \\
&= \int_{-\infty}^{0} h'(u) \int_{-\infty}^{u} f'_{H_\gamma}(v)dvdu + \int_{0}^{+\infty} h'(u) \int_{v}^{+\infty} -f'_{H_\gamma}(v)dvdu \\
&= -\int_{-\infty}^{0}\int_{-\infty}^{0} \mathbb{1}_{\{v\leqslant u\leqslant 0\}} \rho_\gamma(v) f_{H_\gamma}(v) h'(u)dudv + \\
&\qquad \int_{0}^{+\infty}\int_{0}^{+\infty} \mathbb{1}_{\{v\geqslant u\geqslant 0\}} \rho_\gamma(v) f_{H_\gamma}(v) h'(u)dudv \\
&= -\int_{-\infty}^{0}\int_{v}^{0} h'(u)du\, \rho_\gamma(v) f_{H_\gamma}(v)dv + \int_{0}^{+\infty}\int_{0}^{v} h'(u)du\, \rho_\gamma(v) f_{H_\gamma}(v)dv \\
&= \int_{\mathbb{R}} \left(h(v) - h(0)\right) \rho_\gamma(v) f_{H_\gamma}(v)dv = \mathbb{E}\left(h(H_\gamma)\rho_\gamma(H_\gamma)\right) - h(0)\mathbb{E}\left(\rho_\gamma(H_\gamma)\right)
\end{aligned}$$

Now, by symmetry of H_γ and by parity of ρ_γ, we have $\mathbb{E}\left(\rho_\gamma(H_\gamma)\right) = 0$. Hence,

$$\mathbb{E}\left(h'(H_\gamma)\right) = \mathbb{E}\left(\rho_\gamma(H_\gamma)h(H_\gamma)\right) = \mathbb{E}\left(\frac{1}{\gamma^2}\left[H_\gamma - \Psi'\left(\frac{H_\gamma}{\gamma^2}\right)\right] h(H_\gamma)\right)$$

□

Remark 4.6.2. We thus have

$$\begin{aligned}
Y \overset{\mathcal{L}}{=} H_\gamma &\iff \mathbb{E}\left(\mathcal{L}_{\Phi,\gamma} h(Y)\right) = 0 \quad \forall\, h \in \widetilde{\mathcal{H}}_{\Phi,\gamma} \\
&\iff \mathbb{E}\left(h'(Y)\right) = \frac{1}{\gamma^2}\mathbb{E}\left(\left[Y - \psi'\left(\frac{Y}{\gamma^2}\right)\right] h(Y)\right) \quad \forall\, h \in \widetilde{\mathcal{H}}_{\Phi,\gamma}
\end{aligned}$$

The usual caracterisation of the law $\mathscr{N}(0,\gamma^2)$ can be recovered taking $\Phi = 1$ in the last formula, that is $\Psi' = 0$. If we think of γ as a parameter going to $+\infty$, we have a small correction to the Gaussian distribution that takes the form of a small perturbation of the characteristic operator $\mathcal{L}_{\mathscr{N}(0,\gamma^2)}$, i.e. $\mathcal{L}_{\Phi,\gamma} = \mathcal{L}_{\mathscr{N}(0,\gamma^2)} + \psi'(\cdot/\gamma^2)/\gamma^2$.

4.6.2 Perturbation of the Gaussian operator and Edgeworth development

For a random variable X, denote by ϕ_X its Fourier transform

$$\phi_X(u) := \mathbb{E}\left(e^{iuX}\right)$$

Note $\phi_X(u) := \mathcal{F} f_X(u)$ if X admits a Lebesgue-density f_X.

Let $H_\gamma \sim \mathscr{H}(\Phi, \gamma)$, and suppose that $\int_{\mathbb{R}} \left|f_{H_\gamma}(x)\right|^2 dx < \infty$. Denote $\Phi_\gamma := \Phi(\cdot/\gamma)$ and let $(\mathcal{H}_k)_k$ be the normalised Hermite polynomials defined by their Rodrigues form

$$\mathcal{H}_k(y) := e^{y^2/2}\left(-\frac{d}{dy}\right)^k e^{-y^2/2}$$

They satisfy $\mathbb{E}\left(\mathcal{H}_k(G)\mathcal{H}_\ell(G)\right) = h_k \mathbb{1}_{\{k=\ell\}}$ for $G \sim \mathscr{N}(0,1)$ and $h_k > 0$. Thus,

$$\phi_{H_\gamma}(u) := \int_{\mathbb{R}} e^{iux} f_{H_\gamma}(x)dx = \int_{\mathbb{R}} e^{iux}\Phi\left(\frac{x}{\gamma^2}\right) e^{-\frac{1}{2}\left(\frac{x}{\gamma}\right)^2}\frac{dx}{c_\gamma} = \frac{\gamma}{c_\gamma}\int_{\mathbb{R}} \Phi_\gamma(y) e^{iu\gamma y - \frac{y^2}{2}} dy$$

Since $\int_{\mathbb{R}} \left|\Phi_\gamma(x)\right|^2 e^{-x^2} dx = \left(\frac{c_\gamma}{\gamma}\right)^2 \int_{\mathbb{R}} \left|f_{H_\gamma}(x)\right|^2 dx < \infty$, we can write

$$\Phi_\gamma = \sum_{k\geqslant 0} h_k^{-1}\mathbb{E}\left(\mathcal{H}_k(G)\Phi_\gamma(G)\right)\mathcal{H}_k$$

which implies

$$f_{H_\gamma}(\gamma y) = \frac{1}{c_\gamma}\sum_{k\geqslant 0} h_k^{-1}\mathbb{E}\left(\mathcal{H}_k(G)\Phi_\gamma(G)\right)\mathcal{H}_k(y)e^{-y^2/2}$$

A development of the form

$$g_\gamma^{(k)}(y) = e^{-(y/\gamma)^2/2}\sum_{\ell=0}^{k} a_\ell(\gamma)\mathcal{H}_k\left(\frac{y}{\gamma}\right)$$

is said to be an *Edgeworth development* of a random variable. For such a development, truncated at a certain order, there is no possibility to obtain a probability density due to the sign changes of the Hermite polynomials. Without truncation, the function $y \mapsto g_\gamma^{(\infty)}(y)e^{(y/\gamma)^2/2}$ can still be positive and we get a probabilistic penalisation. From this point of view, mod-Gaussian convergence means "infinite Edgeworth development".

4.6.3 Approximation by signed measures

Using the Rodrigues form of the Hermite polynomials, we have

$$f_{H_\gamma}(\gamma y) = c_\gamma^{-1}\sum_{k\geqslant 0} h_k^{-1}\mathbb{E}\left(\mathcal{H}_k(G)\Phi_\gamma(G)\right)\left(-\frac{d}{dy}\right)^k e^{-y^2/2}$$

Taking the Fourier transform and using the fact that $\mathcal{F}(Df)(u) = -iu\,\mathcal{F}f(u)$, we get

$$\begin{aligned}\phi_{H_\gamma}(u) &= c_\gamma^{-1}\sum_{k\geqslant 0} h_k^{-1}\mathbb{E}\left(\mathcal{H}_k(G)\Phi_\gamma(G)\right)(iu/\gamma)^k \mathcal{F}\left(x\mapsto e^{-(x/\gamma)^2/2}\right)(u)\\ &= \gamma c_\gamma^{-1}\sqrt{2\pi}e^{-\gamma^2u^2/2}\sum_{k\geqslant 0} h_k^{-1}\mathbb{E}\left(\mathcal{H}_k(G)\Phi_\gamma(G)\right)(iu/\gamma)^k\end{aligned}$$

the last development being convergent in $L^2(e^{-x^2}dx)$ since Φ_γ belongs to this space. As $\phi_{\gamma G}(u) = e^{-\gamma^2u^2/2}$, denoting by

$$\widetilde{\Phi}_\gamma(u) := \gamma\sqrt{2\pi}\sum_{k\geqslant 0} h_k^{-1}\mathbb{E}\left(\mathcal{H}_k(G)\Phi_\gamma(G)\right)(iu/\gamma)^k$$

we get

$$\phi_{H_\gamma}(u) = \widetilde{\Phi}_\gamma(u)\phi_{\gamma G}(u) \tag{4.25}$$

In particular, if we know that $\Phi(u) = \lim_{\gamma\to\infty}\phi_{H_\gamma}(u)/\phi_{\gamma G}(u)$ exists (the limit being locally uniform in u), then, we have

$$\Phi(u) = \lim_{\gamma\to\infty}\widetilde{\Phi}_\gamma(u) \tag{4.26}$$

i.e. $\widetilde{\Phi}_\gamma$ is an approximation of Φ.

The construction of a random variable whose distribution satisfies equations (4.25) and (4.26) is not always possible. In the flavour of [4], one can be interested in a signed measure approximation of sequences converging in the mod-Gaussian sense. A case of interest is the following : suppose that $\widetilde{\Phi}_\gamma$ can be approximated by

$$\Phi_\gamma^\sharp(u) := e^{P(i\theta)}$$

P being a polynomial satisfying the symmetry condition of theorem 3.4.7, that is $P(i\theta) = P(\theta)$ for all $\theta\in\mathbb{R}$, and such that $P(0) = 0$. Let $\mu_\gamma := \mathcal{F}^{-1}\left(\Phi_\gamma^\sharp\,\phi_{\gamma G}\right)$ be the signed measure obtained by inverting equation (4.25). Then, a Stein operator of μ_γ is given by

$$\mathcal{L}_{\Phi^\sharp,\gamma} := \mathcal{L}_{\mathcal{N}(0,\gamma^2)} - P'\left(-\frac{d}{dx}\right)$$

Indeed, suppose $\mathcal{F}\mu_\gamma\in L^1$ with density $f\in\mathcal{C}^1$. Such an operator satisfies $\int_\mathbb{R}\mathcal{L}_{\Phi^\sharp,\gamma}g\cdot f = 0$ for all functions g of class $\mathcal{C}^\infty\cap L^1$ that vanish at $\pm\infty$, which amounts by integration by parts to the following equation

$$\mathcal{L}^*_{\Phi^\sharp,\gamma}f(x) := xf(x) + \gamma^2 f'(x) - P'\left(\frac{d}{dx}\right)f(x) = 0$$

Taking the Fourier transform, setting $\hat{f} := \mathcal{F}f$ and using the relations $\mathcal{F}\left(x\mapsto xf(x)\right)(\xi) = -i\frac{d}{d\xi}\hat{f}(\xi)$ and $\mathcal{F}\left(f'\right)(\xi) = (-i\xi)\hat{f}(\xi)$, we get

$$-i\frac{d}{d\xi}\hat{f}(\xi) + \left(\gamma^2(-i\xi) - P'(i\xi)\right)\hat{f}(\xi) = 0$$

The integration of this equation gives

$$\hat{f}(\xi) = \hat{f}(0) \exp\left(-\gamma^2 \frac{\xi^2}{2} + P(i\xi) - P(0)\right) = \hat{f}(0) \exp\left(-\gamma^2 \frac{\xi^2}{2} + P(\xi)\right)$$

By Fourier inversion,

$$f(x) = \hat{f}(0) \int_{\mathbb{R}} \exp\left(-i\xi x - \gamma^2 \frac{\xi^2}{2} + P(\xi)\right) \frac{d\xi}{2\pi}$$

which is (proportional to) the density of the measure μ_γ.

In this setting, Stein's method with an operator such as $\mathcal{L}_{\Phi^\sharp,\gamma}$ allows to approximate $\mathbb{P}_{H_\gamma}$ (hence the sequence of non-renormalised random variables) with μ_γ, which is to relate to [4] where Kolmogorov approximations (in the Poisson setting) were found with respect to a signed measure, and to [3] where this type of correction to $\mathcal{L}_{\mathcal{N}(0,\gamma^2)}$ is discussed in details (see also [84]). In [3], a perturbation of $\mathcal{L}_{\mathcal{N}(0,1)}$ is done with a polynomial in the operator of differentiation, and this polynomial is the truncation of the cumulant generating series, which is the essence of mod-* convergence since the limiting function Φ is obtained by a suitable renormalisation of the cumulant function.

Example 4.6.3. In the case of example 3.4.8, as an approximation of $\widetilde{\Phi}_\gamma$ is $\Phi_\gamma^\sharp = \Phi(\cdot/\gamma^2) : x \mapsto e^{-Cx^4/(4\gamma^8)}$, the condition is fullfilled and one can have an Edgeworth development by using a suitable truncation of the Taylor development of $e^{-Cx^4/4}$ in addition to a signed measure approximation of density $x \mapsto \int_{\mathbb{R}} \exp\left(-i\xi x - \gamma^2 \frac{\xi^2}{2} - C\xi^4/(4\gamma^8)\right) \frac{d\xi}{2\pi}$.

4.7 The sum of i.i.d. symmetric random variables

We develop the important example of the sum of i.i.d. random variables. In order to fit with theorem 3.4.7, we only consider the symmetric case.

4.7.1 A mod-Gaussian approximation theorem

Theorem 4.7.1 (Mod-Gaussian bounds for the sum of i.i.d. random variables)**.** *Let $(X_k)_k$ be a sequence of i.i.d. symmetric random variables having the same law as X such that $\mathbb{E}(X) = 0$, $\mathbb{E}(X^2) = 1$ and $\mathbb{E}(X^4) < 3$. Set*

$$\begin{aligned}
W_n &:= \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \\
\gamma_n &:= n^{1/4} \\
Z_n &:= \gamma_n W_n \\
C &:= \frac{3 - \mathbb{E}(X^4)}{6} > 0 \\
\Phi_C(x) &:= e^{-C\frac{x^4}{4}}
\end{aligned}$$

Let $H_n \sim \mathscr{H}(\Phi_C, \gamma_n)$. Then, for $c_1 := \sqrt{\pi/2}\mathbb{E}\left(\Phi_C(G)\right)$ with $G \sim \mathscr{N}(0,1)$

$$|\mathbb{E}\left(h(Z_n)\right) - \mathbb{E}\left(h(H_n)\right)| \leqslant \frac{4\sqrt{2(1-C)}}{\gamma_n} \left\|h'\right\|_\infty + \frac{4}{\gamma_n^2} \left\|h\right\|_\infty \left(\mathbb{E}\left(|X|\right) \vee \mathbb{E}\left(\frac{|X|^3}{2}\right) c_1 C + \frac{1}{\gamma_n^2}\right)$$

In particular, setting

$$\mathcal{H}_\Phi := \left\{ h \in \mathcal{C}^1(\mathbb{R}) / \left\|h\right\|_\infty \leqslant 1, \left\|h'\right\|_\infty \leqslant 1, \lim_{\pm\infty} h = 0, \int_{\mathbb{R}} \left|h'\right| < \infty \right\}$$

$$d_{\mathcal{H}_\Phi}(X,Y) := \sup_{h \in \mathcal{H}_\Phi} |\mathbb{E}\left(h(X)\right) - \mathbb{E}\left(h(Y)\right)|$$

we have, for n large enough

$$d_{\mathcal{H}_\Phi}(Z_n, H_n) \leqslant \frac{4\sqrt{2}}{\gamma_n}$$

Proof. We start with the usual Stein's argument : for $h \in \mathcal{H}_\Phi$

$$h_{H_n}(x) := h(x) - \mathbb{E}\left(h(H_n)\right) = \mathcal{L}_{H_n} \mathcal{L}_{H_n}^{-1} h_{H_n}(x)$$

Setting $g := \mathcal{L}_{H_n}^{-1} h_{H_n}$, we get

$$\mathbb{E}\left(h(Z_n)\right) - \mathbb{E}\left(h(H_n)\right) = \mathbb{E}\left(\mathcal{L}_{H_n} g(Z_n)\right) = \mathbb{E}\left(g'(Z_n) - \rho_{\gamma_n}(Z_n) g(Z_n)\right)$$

We recall that

$$\rho_{\gamma_n}(x) := \frac{1}{\gamma_n^2}\left(x - (\log \Phi_C)'\left(\frac{x}{\gamma_n^2}\right)\right) = \frac{x}{\gamma_n^2} + C\frac{x^3}{\gamma_n^8}$$

For the usual Stein's operator given by $\mathcal{L}_{\mathscr{N}(0,\gamma_n^2)} f(x) := f'(x) - (x/\gamma_n^2) f(x)$, we have

$$\mathbb{E}\left(\mathcal{L}_{\mathscr{N}(0,\gamma_n^2)} g(Z_n)\right) = \mathbb{E}\left(g'(Z_n) - \frac{Z_n}{\gamma_n^2} g(Z_n)\right) = \mathbb{E}\left(g'(Z_n) - g'\left(Z_n^{(0)}\right)\right)$$

where $Z_n^{(0)}$ is the zero-bias transform of Z_n. By (4.19), we have

$$Z_n^{(0)} \stackrel{\mathcal{L}}{=} Z_n + \frac{\gamma_n}{\sqrt{n}}\left(X_I^{(0)} - X_I\right) = Z_n + \frac{1}{\gamma_n}\left(X_I^{(0)} - X_I\right)$$

where $I \sim \mathscr{U}\left([\![1,n]\!]\right)$ is a random variable independent of $(X_k)_k$ and $(X_k^{(0)})_k$, those two last sequences being independent, and $(X_k^{(0)})_k$ being a sequence of i.i.d. random variables distributed according to the zero-bias distribution of X.

In particular,

$$\begin{aligned}\left|\mathbb{E}\left(\mathcal{L}_{\mathscr{N}(0,\gamma_n^2)} g(Z_n)\right)\right| &\leqslant \mathbb{E}\left(\left|g'(Z_n) - g'\left(Z_n^{(0)}\right)\right|\right) \\ &\leqslant \left\|g''\right\|_\infty \mathbb{E}\left(\left|Z_n - Z_n^{(0)}\right|\right) \\ &= \frac{1}{\gamma_n}\left\|g''\right\|_\infty \mathbb{E}\left(\left|X_I^{(0)} - X_I\right|\right)\end{aligned}$$

Moreover, by independence and the i.i.d. property,

$$\mathbb{E}\left(\left|X_I^{(0)} - X_I\right|\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(\left|X_k^{(0)} - X_k\right|\right) = \mathbb{E}\left(|X^{(0)} - X|\right)$$

For the perturbative operator, we get

$$\mathbb{E}\left((\mathcal{L}_{H_n} - \mathcal{L}_{\mathcal{N}(0,\gamma_n^2)})g(Z_n)\right) = -\frac{C}{\gamma_n^8}\mathbb{E}\left(Z_n^3 g(Z_n)\right) =: -\frac{C}{\gamma_n^8}\mathbb{E}\left(Z_n\tilde{g}(Z_n)\right)$$

with $\tilde{g}(x) := x^2 g(x)$. Using the same technique, we write

$$\mathbb{E}\left(Z_n\tilde{g}(Z_n)\right) = \gamma_n^2\mathbb{E}\left(\tilde{g}'\left(Z_n^{(0)}\right)\right) = \gamma_n^2\mathbb{E}\left(2Z_n^{(0)}g\left(Z_n^{(0)}\right) + \left(Z_n^{(0)}\right)^2 g'\left(Z_n^{(0)}\right)\right)$$

Hence,

$$\begin{aligned}
\left|\mathbb{E}\left((\mathcal{L}_{H_n} - \mathcal{L}_{\mathcal{N}(0,\gamma_n^2)})g(Z_n)\right)\right| &\leqslant \frac{C}{\gamma_n^8}\left(2\gamma_n^2\mathbb{E}\left(\left|Z_n^{(0)}\right|\right)\|g\|_\infty + \gamma_n^2\mathbb{E}\left(\left|Z_n^{(0)}\right|^2\right)\|g'\|_\infty\right) \\
&\leqslant \frac{C}{\gamma_n^6}\Bigg(2\|g\|_\infty\frac{\gamma_n}{\sqrt{n}}\left[(n-1)\mathbb{E}\left(|X|\right) + \mathbb{E}\left(\left|X^{(0)}\right|\right)\right] \\
&\qquad + \|g'\|_\infty\left(\frac{\gamma_n}{\sqrt{n}}\right)^2\left[(n-1)\mathbb{E}\left(X^2\right) + \mathbb{E}\left(\left|X^{(0)}\right|^2\right)\right]\Bigg) \\
&\leqslant \frac{C}{\gamma_n^6}\Bigg(2\|g\|_\infty\frac{n}{\gamma_n}\mathbb{E}\left(|X|\right)\vee\mathbb{E}\left(\left|X^{(0)}\right|\right) \\
&\qquad + \frac{n}{\gamma_n^2}\|g'\|_\infty\mathbb{E}\left(|X|^2\right)\vee\mathbb{E}\left(\left|X^{(0)}\right|^2\right)\Bigg)
\end{aligned}$$

Using (4.18) for well-choosen functions, we can prove that

$$\begin{aligned}
\mathbb{E}\left(\left|X^{(0)}\right|\right) &= \frac{\mathbb{E}\left(|X|^3\right)}{2\mathbb{E}\left(X^2\right)} \\
\mathbb{E}\left(\left(X^{(0)}\right)^2\right) &= \frac{\mathbb{E}\left(X^4\right)}{3\mathbb{E}\left(X^2\right)}
\end{aligned}$$

As $\gamma_n^4 = n$, $\mathbb{E}\left(X^2\right) = 1$, and $\mathbb{E}\left(X^4\right) := 3(1-2C) < 3$, we have by independence of X and $X^{(0)}$

$$\mathbb{E}\left(\left|X - X^{(0)}\right|\right) \leqslant \sqrt{\mathbb{E}\left(\left|X - X^{(0)}\right|^2\right)} = \sqrt{\mathbb{E}\left(X^2 + (X^{(0)})^2\right)} = \sqrt{2(1-C)}$$

and finally

$$\left|\mathbb{E}\left(\mathcal{L}_{H_n}g(Z_n)\right)\right| \leqslant \frac{\sqrt{2(1-C)}}{\gamma_n}\|g''\|_\infty + \frac{2C}{\gamma_n^3}\|g\|_\infty\mathbb{E}\left(|X|\right)\vee\mathbb{E}\left(\frac{|X|^3}{2}\right) + \frac{C}{\gamma_n^4}\|g'\|_\infty$$

The last part of Stein's method consists in using the estimates for $\left\|D^k g\right\|_\infty = \left\|D^k \mathcal{L}_{H_n}^{-1} h_{H_n}\right\|_\infty$ with $k \in \{0,1,2\}$, which was done in the lemma 4.8.5. Pluging the inequalities (4.44) to (4.46) in the last inequality, we get

$$|\mathbb{E}(h(Z_n)) - \mathbb{E}(h(H_n))| \leqslant \frac{4\sqrt{2(1-C)}}{\gamma_n} \left\|h'\right\|_\infty + \frac{2}{\gamma_n^2} \|h_{H_n}\|_\infty \left(\mathbb{E}(|X|) \vee \mathbb{E}\left(\frac{|X|^3}{2}\right) c_1 C + \frac{1}{\gamma_n^2}\right) \quad (4.27)$$

Using $\|h_{H_n}\|_\infty \leqslant 2\|h\|_\infty$, we get the desired bound.

□

Remark 4.7.2. With the suitable rescaling of Z_n and H_n, we get

$$\left|\mathbb{E}\left(h\left(\frac{Z_n}{\gamma_n}\right)\right) - \mathbb{E}\left(h\left(\frac{H_n}{\gamma_n}\right)\right)\right| \leqslant \frac{1}{\gamma_n^2}\left(4\sqrt{2(1-C)}\left\|h'\right\|_\infty + \|h\|_\infty O(1)\right) = O\left(\frac{1}{\sqrt{n}}\right)$$

which corresponds to no improvement of the CLT. This bound will be improved in theorem 4.7.4 conditionally to a conjectural estimate.

4.7.2 A Kolmogorov approximation

Following the steps of Stein ([94] p. 36), we have the

Corollary 4.7.3 (Kolmogorov bounds)**.** *Let $(X_k)_k$ satisfying the hypothesis of theorem 4.7.1. Then*

$$d_{\mathrm{Kol}}(Z_n, H_n) := \sup_{x\in\mathbb{R}} |\mathbb{P}(Z_n \leqslant x) - \mathbb{P}(H_n \leqslant x)| \leqslant \frac{4((1-C)/\pi)^{1/4}}{\gamma_n} + O\left(\frac{1}{\gamma_n^2}\right) \quad (4.28)$$

Proof. Set

$$h_{x,\delta}(y) := \mathbb{1}_{\{y\leqslant x\}} + \left(1 - \frac{y-x}{\delta}\right)\mathbb{1}_{\{x\leqslant y\leqslant x+\delta\}}$$

For all x, we have $h_{x-\delta,\delta}(y) \leqslant \mathbb{1}_{\{y\leqslant x\}} \leqslant h_{x,\delta}(y) \leqslant \mathbb{1}_{\{y\leqslant x+\delta\}}$, which implies that

$$\mathbb{E}(h_{x-\delta,\delta}(Z_n)) \leqslant \mathbb{P}(Z_n \leqslant x) \leqslant \mathbb{E}(h_{x,\delta}(Z_n)) \quad (4.29)$$

Moreover, we have

$$\begin{aligned} \|h_{x,\delta} - \mathbb{E}(h_{x,\delta}(H_n))\|_\infty &\leqslant 1 \\ \left\|h'_{x,\delta}\right\|_\infty &\leqslant \frac{1}{\delta} \end{aligned}$$

Using those inequalities and (4.27), and setting $\alpha := \mathbb{E}(|X|) \vee \mathbb{E}\left(\frac{|X|^3}{2}\right) c_1 C + 1$, we get

$$|\mathbb{E}(h_{x,\delta}(Z_n)) - \mathbb{E}(h_{x,\delta}(H_n))| \leqslant \frac{4\sqrt{2(1-C)}}{\gamma_n}\frac{1}{\delta} + \frac{2\alpha}{\gamma_n^2}$$

By (4.29), we have

$$\begin{aligned}
\mathbb{P}(Z_n \leqslant x) &\leqslant \mathbb{E}(h_{x,\delta}(Z_n)) \\
&\leqslant \mathbb{E}(h_{x,\delta}(H_n)) + \frac{4\sqrt{2(1-C)}}{\gamma_n \delta} + \frac{2\alpha}{\gamma_n^2} \\
&= \mathbb{P}(H_n \leqslant x) + \mathbb{E}\left(\left(1 - \frac{H_n - x}{\delta}\right) \mathbb{1}_{\left\{0 \leqslant \frac{H_n - x}{\delta} \leqslant 1\right\}}\right) + \frac{4\sqrt{2(1-C)}}{\gamma_n \delta} + \frac{2\alpha}{\gamma_n^2} \\
&\leqslant \mathbb{P}(H_n \leqslant x) + \mathbb{P}(0 \leqslant H_n - x \leqslant \delta) + \frac{4\sqrt{2(1-C)}}{\gamma_n \delta} + \frac{2\alpha}{\gamma_n^2} \\
&\leqslant \mathbb{P}(H_n \leqslant x) + \frac{\delta}{c_{\gamma_n}} + \frac{4\sqrt{2(1-C)}}{\gamma_n \delta} + \frac{2\alpha}{\gamma_n^2}
\end{aligned}$$

This last inequality comes from the fact that

$$\mathbb{P}(0 \leqslant H_\gamma - x \leqslant \delta) = \int_0^\delta e^{-P_{q_\gamma}(y-x)} \frac{dy}{c_\gamma} \leqslant \frac{\delta}{c_\gamma} \sup_{y \in \mathbb{R}_+} \left\{ e^{-P_{q_\gamma}(y-x)} \right\} = \frac{\delta}{c_\gamma}$$

Optimising in δ the LHS of the former inequality gives

$$\delta = \sqrt{c_{\gamma_n} \frac{4\sqrt{2(1-C)}}{\gamma_n}}$$

and the optimal value

$$\mathbb{P}(Z_n \leqslant x) - \mathbb{P}(H_n \leqslant x) \leqslant 2\sqrt{\frac{4\sqrt{2(1-C)}}{c_{\gamma_n}\gamma_n}} + \frac{2\alpha}{\gamma_n^2}$$

As by (3.20), $c_\gamma = \gamma\sqrt{2\pi}\mathbb{E}\left(e^{-C\frac{G^4}{4\gamma^4}}\right) \leqslant \gamma\sqrt{2\pi}$, we finally have

$$\mathbb{P}(Z_n \leqslant x) - \mathbb{P}(H_n \leqslant x) \leqslant \frac{4((1-C)/\pi)^{1/4}}{\gamma_n} + \frac{2\alpha}{\gamma_n^2}$$

The corresponding lower bound follows from the same manipulations using the lower bound in (4.29).

□

4.7.3 A conjectural mod-Gaussian approximation theorem

As noticed in remark 4.7.2, the bound (4.27) is not optimal since a suitable rescaling gives the same speed of convergence as the usual CLT. We nevertheless have the following improvement, conditional to an estimate that is conjectured to be true.

Theorem 4.7.4 (Conditional mod-Gaussian bounds for the sum of i.i.d. random variables)**.** *Let $(X_k)_k$ be a sequence satisfying the hypotheses of theorem 4.7.1. Suppose moreover the following conjecture : there exists a constant $A > 0$ independent of n such that*

$$\left\| D^3 \mathcal{L}_{H_n}^{-1} h_{H_n} \right\|_\infty \leqslant A \left\| D^2 h \right\|_\infty$$

Then, we have the following bound

$$\left|\mathbb{E}\left(h(Z_n)\right) - \mathbb{E}\left(h(H_n)\right)\right| \leqslant A\frac{2-3C}{\gamma_n^2}\left\|h''\right\|_\infty + \frac{12\,C}{\gamma_n^4}\left\|h\right\|_\infty$$

In particular, for n large enough

$$d_{\mathcal{H}_\Phi}(Z_n, H_n) \leqslant \frac{A(2-3C)}{\gamma_n^2}$$

Proof. For $h \in \mathcal{H}_\Phi$ and $g := \mathcal{L}_{H_n}^{-1} h_{H_n}$, we get

$$\mathbb{E}\left(h(Z_n)\right) - \mathbb{E}\left(h(H_n)\right) = \mathbb{E}\left(\mathcal{L}_{H_n} g(Z_n)\right) = \mathbb{E}\left(g'(Z_n) - \rho_{\gamma_n}(Z_n) g(Z_n)\right)$$

We treat the case of the Stein's operator $\mathcal{L}_{\mathcal{N}(0,\gamma_n^2)}$ like before, with

$$\mathbb{E}\left(\mathcal{L}_{\mathcal{N}(0,\gamma_n^2)} g(Z_n)\right) = \mathbb{E}\left(g'(Z_n) - \frac{Z_n}{\gamma_n^2} g(Z_n)\right) = \mathbb{E}\left(g'(Z_n) - g'\left(Z_n^{(0)}\right)\right)$$

Setting

$$\Delta Z_n := Z_n^{(0)} - Z_n = \frac{X_I^{(0)} - X_I}{\gamma_n}$$

we have

$$\begin{aligned} g'(Z_n^{(0)}) - g'(Z_n) - \Delta Z_n g''(Z_n) &= \int_0^1 g''\left(Z_n + w\Delta Z_n\right)\Delta Z_n \, dw - \Delta Z_n g''(Z_n) \\ &= \Delta Z_n \int_0^1 \int_0^1 g'''\left(Z_n + wu\Delta Z_n\right) w \Delta Z_n du \, dw \\ &= \frac{(\Delta Z_n)^2}{2} \int_0^1 \int_0^1 g'''\left(Z_n + u\sqrt{v}\Delta Z_n\right) du \, dv \end{aligned}$$

By independence of the terms and since $\mathbb{E}\left(X_I^{(0)}\right) = 0$ as seen in (4.18) taking $f(x) = x^2$, we have

$$\begin{aligned} \mathbb{E}\left(\Delta Z_n g''(Z_n)\right) &= \frac{1}{\gamma_n}\mathbb{E}\left((X_I^{(0)} - X_I)\, g''\left(\frac{1}{\gamma_n}\sum_{k=1}^n X_k\right)\right) \\ &= \frac{1}{\gamma_n}\mathbb{E}\left(X_I^{(0)}\right)\mathbb{E}\left(g''\left(\sum_{k=1}^n X_k\right)\right) - \frac{1}{\gamma_n}\mathbb{E}\left(X_I g''\left(\frac{1}{\gamma_n}\sum_{k=1}^n X_k\right)\right) \\ &= -\frac{1}{\gamma_n}\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^n X_k g''\left(\frac{1}{\gamma_n}\sum_{k=1}^n X_k\right)\right) \end{aligned}$$

by integrating on I which is independent of $(X_k)_k$. As $n = \gamma_n^4$, we hence have

$$\mathbb{E}\left(\Delta Z_n g''(Z_n)\right) = -\frac{1}{\gamma_n^4}\mathbb{E}\left(S_n g''(S_n)\right) = -\frac{1}{\gamma_n^4}\mathbb{E}\left(S_n^2\right)\mathbb{E}\left(g'''(Z_n^{(0)})\right) = -\frac{1}{\gamma_n^2}\mathbb{E}\left(g'''(Z_n^{(0)})\right)$$

Let $U, V \sim \mathscr{U}([0,1])$ be two independent random variables. Then,

$$\begin{aligned}\left|\mathbb{E}\left(\mathcal{L}_{\mathcal{N}(0,\gamma_n^2)} g(Z_n)\right)\right| &= \frac{1}{\gamma_n^2}\left|\mathbb{E}\left(\frac{(X_I - X^{(0)})^2}{2} g'''\left(Z_n + 2V\sqrt{U}\Delta Z_n\right) - g'''\left(Z_n^{(0)}\right)\right)\right| \\ &\leqslant \frac{1}{\gamma_n^2}\|g'''\|_\infty \left(\mathbb{E}\left(\frac{(X_I - X_I^{(0)})^2}{2}\right) + 1\right)\end{aligned}$$

And by independence of X and $X^{(0)}$ and by the i.i.d. property

$$\mathbb{E}\left((X_I - X_I^{(0)})^2\right) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left((X_k - X_k^{(0)})^2\right) = \mathbb{E}\left((X - X^{(0)})^2\right) = \mathbb{E}\left(X^2 + (X^{(0)})^2\right) = 2(1-C)$$

which finally gives

$$\left|\mathbb{E}\left(\mathcal{L}_{\mathcal{N}(0,\gamma_n^2)} g(Z_n)\right)\right| \leqslant \frac{2-C}{\gamma_n^2}\|g'''\|_\infty$$

For the perturbative operator, we get

$$\mathbb{E}\left((\mathcal{L}_{H_n} - \mathcal{L}_{\mathcal{N}(0,\gamma_n^2)}) g(Z_n)\right) = -\frac{C}{\gamma_n^8}\mathbb{E}\left(Z_n^3 g(Z_n)\right) =: -\frac{C}{\gamma_n^8}\mathbb{E}\left(Z_n \tilde{g}(Z_n)\right) \quad \text{with} \quad \tilde{g}(x) := x^2 g(x)$$

Using iteratively the 0-bias transform, we have

$$\begin{aligned}\mathbb{E}\left(Z_n \tilde{g}(Z_n)\right) &= \gamma_n^2 \mathbb{E}\left(\tilde{g}'\left(Z_n^{(0)}\right)\right) \\ &= \gamma_n^2 \mathbb{E}\left(2Z_n^{(0)} g\left(Z_n^{(0)}\right) + \left(Z_n^{(0)}\right)^2 g'\left(Z_n^{(0)}\right)\right) \\ &= \gamma_n^2 \mathbb{E}\left(2\mathbb{E}\left(\left(Z_n^{(0)}\right)^2\right) g'\left(Z_n^{(0,0)}\right) + \left(Z_n^{(0)}\right)^2 g'\left(Z_n^{(0)}\right)\right)\end{aligned}$$

where $Z_n^{(0,0)}$ is the 0-bias transform of $Z_n^{(0)}$, what we can do since 0-biaising preserve the property of being of expectation 0 for symmetric random variables, as seen in (4.18) taking $f(x) = x^2$.

As $Z_n^{(0)} \stackrel{\mathcal{L}}{=} \frac{1}{\gamma_n}\left(\sum_{k\neq I} X_k + X_I^{(0)}\right)$ and $\mathbb{E}\left(\left(X^{(0)}\right)^2\right) = 1 - 2C$, we have

$$\mathbb{E}\left(\left(Z_n^{(0)}\right)^2\right) = \frac{1}{\gamma_n^2}\left((n-1)\mathbb{E}\left(X^2\right) + \mathbb{E}\left(\left(X^{(0)}\right)^2\right)\right) = \frac{1}{\gamma_n^2}\left(\gamma_n^4 - 2C\right)$$

which gives

$$\left|\mathbb{E}\left((\mathcal{L}_{H_n} - \mathcal{L}_{\mathcal{N}(0,\gamma_n^2)}) g(Z_n)\right)\right| \leqslant \frac{C}{\gamma_n^8} 3\gamma_n^2 \mathbb{E}\left(\left(Z_n^{(0)}\right)^2\right)\|g'\|_\infty = 3C\frac{\gamma_n^4 - 2C}{\gamma_n^8}\|g'\|_\infty \leqslant \frac{3C}{\gamma_n^4}\|g'\|_\infty$$

Finally

$$\left|\mathbb{E}\left(\mathcal{L}_{H_n} g(Z_n)\right)\right| \leqslant \frac{2-C}{\gamma_n^2}\|g'''\|_\infty + \frac{3C}{\gamma_n^4}\|g'\|_\infty$$

To conclude with Stein's methodology, we need the estimates $\left\|D^k g\right\|_\infty = \left\|D^k \mathcal{L}_{H_n}^{-1} h_{H_n}\right\|_\infty$ with $k \in \{1,3\}$. Note that the estimate for $k=3$ is a conjecture. Using these results, we have

$$|\mathbb{E}\left(h(Z_n)\right) - \mathbb{E}\left(h(H_n)\right)| \leqslant A\frac{2-3C}{\gamma_n^2}\left\|h''\right\|_\infty + \frac{6C}{\gamma_n^4}\left\|h_{H_n}\right\|_\infty \tag{4.30}$$

Last, using $\left\|h_{H_n}\right\|_\infty \leqslant 2\left\|h\right\|_\infty$, we have the bound.

□

Remark 4.7.5. With the suitable rescaling of Z_n and H_n, we get

$$\left|\mathbb{E}\left(h\left(\frac{Z_n}{\gamma_n}\right)\right) - \mathbb{E}\left(h\left(\frac{H_n}{\gamma_n}\right)\right)\right| \leqslant \frac{1}{\gamma_n^4}\left(A(2-3C)\left\|h''\right\|_\infty + 12C\left\|h\right\|_\infty\right) = O\left(\frac{1}{n}\right)$$

which corresponds to an improvement of the CLT. This can be understood as an additive correction to the usual norm by writing

$$\left|\mathbb{E}\left(h\left(\frac{Z_n}{\gamma_n}\right)\right) - \mathbb{E}\left(h\left(G\right)\right) + \mathbb{E}\left(h\left(G\right) - h\left(\frac{H_n}{\gamma_n}\right)\right)\right| = \left|\mathbb{E}\left(h\left(\frac{Z_n}{\gamma_n}\right)\right) - \mathbb{E}\left(h\left(G\right)\right) + \mathbb{E}\left(\Delta_n h'\left(K_n\right)\right)\right|$$

where, for $U \sim \mathcal{U}([0,1])$ independent of G, $\Delta_n := G - H_n/\gamma_n$ and

$$\mathbb{E}\left(h\left(G\right) - h\left(\frac{H_n}{\gamma_n}\right)\right) = \mathbb{E}\left(\int_{H_n/\gamma_n}^{G} h'(x)dx\right) = \mathbb{E}\left(h'\left(H_n/\gamma_n + U\Delta_n\right)\Delta_n\right)$$

The search for an additional correction that would give a faster approximation was also developed in [35]. A comparison between the two corrective terms would be interesting.

4.7.4 A conditional Kolmogorov approximation

Conditionaly to (4.30), we have the following

Corollary 4.7.6 (Kolmogorov bounds)**.** *Let $(X_k)_k$ satisfying the hypothesis of theorem 4.7.1. Then*

$$d_{\mathrm{Kol}}(Z_n, H_n) := \sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(Z_n \leqslant x\right) - \mathbb{P}\left(H_n \leqslant x\right)\right| \leqslant \frac{\beta}{\gamma_n^{4/3}} + O\left(\frac{1}{\gamma_n^2}\right) \tag{4.31}$$

with $\beta := (2^{1/3} + 2^{-2/3})((A(2-3C))^{1/4}$.

Proof. Set

$$\begin{aligned} Q(t) &:= 2t^3 - 3t^2 + 1 \\ D(t) &:= \mathbb{1}_{\{t\leqslant 0\}} + Q(t)\mathbb{1}_{\{0\leqslant t\leqslant 1\}} \\ h_{x,\delta}(y) &:= D\left(\frac{y-x}{\delta}\right) \end{aligned}$$

For all x, we have $h_{x-\delta,\delta}(y) \leqslant \mathbb{1}_{\{y\leqslant x\}} \leqslant h_{x,\delta}(y)$, which implies that

$$\mathbb{E}\left(h_{x-\delta,\delta}(Z_n)\right) \leqslant \mathbb{P}\left(Z_n \leqslant x\right) \leqslant \mathbb{E}\left(h_{x,\delta}(Z_n)\right) \tag{4.32}$$

Moreover, we have

$$\left\| h_{x,\delta} - \mathbb{E}\left(h_{x,\delta}(H_n)\right)\right\|_\infty \leqslant 1$$
$$\left\| h''_{x,\delta}\right\|_\infty \leqslant \frac{1}{\delta^2}$$

Using these inequalities and (4.30), and setting $\alpha := A(2-3C)$, we get

$$\left|\mathbb{E}\left(h_{x,\delta}(Z_n)\right) - \mathbb{E}\left(h_{x,\delta}(H_n)\right)\right| \leqslant \frac{\alpha}{\gamma_n^2}\frac{1}{\delta^2} + \frac{6C}{\gamma_n^4}$$

By (4.32), we have

$$\begin{aligned}
\mathbb{P}\left(Z_n \leqslant x\right) &\leqslant \mathbb{E}\left(h_{x,\delta}(Z_n)\right) \\
&\leqslant \mathbb{E}\left(h_{x,\delta}(H_n)\right) + \frac{\alpha}{\gamma_n^2\delta^2} + \frac{6C}{\gamma_n^4} \\
&= \mathbb{P}\left(H_n \leqslant x\right) + \mathbb{E}\left(Q\left(\frac{H_n - x}{\delta}\right)\mathbb{1}_{\left\{0 \leqslant \frac{H_n - x}{\delta} \leqslant 1\right\}}\right) + \frac{\alpha}{\gamma_n^2\delta^2} + \frac{6C}{\gamma_n^4} \\
&\leqslant \mathbb{P}\left(H_n \leqslant x\right) + \mathbb{P}\left(0 \leqslant H_n - x \leqslant \delta\right) + \frac{\alpha}{\gamma_n^2\delta^2} + \frac{6C}{\gamma_n^4} \\
&\leqslant \mathbb{P}\left(H_n \leqslant x\right) + \frac{\delta}{\gamma_n} + \frac{\alpha}{\gamma_n^2\delta^2} + \frac{6C}{\gamma_n^4}
\end{aligned}$$

Optimising in δ the LHS of the former inequality gives

$$\delta = \left(\frac{2\alpha}{\gamma_n}\right)^{1/3}$$

and the optimal value

$$\mathbb{P}\left(Z_n \leqslant x\right) - \mathbb{P}\left(H_n \leqslant x\right) \leqslant (2^{1/3} + 2^{-2/3})\left(\frac{\alpha}{\gamma_n^4}\right)^{1/3} + \frac{6C}{\gamma_n^2}$$

Finally, we have

$$\mathbb{P}\left(Z_n \leqslant x\right) - \mathbb{P}\left(H_n \leqslant x\right) \leqslant (2^{1/3} + 2^{-2/3})\frac{\alpha^{1/3}}{\gamma_n^{4/3}} + \frac{6C}{\gamma_n^2}$$

The corresponding lower bound follows from the same manipulations.

□

Remark 4.7.7. We see that the zero-bias transform is not characteristic of the distribution $H_\gamma \sim \mathscr{H}(\Phi_C, \gamma)$. Indeed, the Stein's equation (4.3) characteristic of the Gaussian distribution is equivalent to the fixed point equation in distribution

$$X \sim \mathscr{N}(0,1) \iff X \stackrel{\mathcal{L}}{=} X^{(0)}$$

but we do not characterise the distribution $\mathscr{H}(\Phi_C, \gamma)$ with such a transformation.

A natural transformation would be the following C-bias transform, defined for a random variable W such that $\mathbb{E}(W) = \mathbb{E}(W^3) = 0$, $\mathbb{E}(W^2) = \gamma^2$ and $\mathbb{E}(W^4) < \infty$ and for all absolutely continuous functions f satisfying $\mathbb{E}\left(\left|W^3 f(W)\right|\right) < \infty$ by

$$\mathbb{E}\left(f'\left(W^{(C)}\right)\right) = \left(\gamma^2 + \frac{4C}{\gamma^6}\mathbb{E}\left(W^4\right)\right)^{-1} \mathbb{E}\left(\rho_C(W) f(W)\right) \tag{4.33}$$

with

$$\rho_C(x) := x + \frac{4C}{\gamma^6} x^3$$

The distribution of $W^{(C)}$ is absolutely continuous with respect to Lebesgue measure and has for density

$$f_{W^{(C)}}(x) = \mathbb{E}\left(\rho_C(W) \mathbb{1}_{\{W \geqslant x\}}\right)$$

The proof of such a result is the same as in the case of the zero-bias transform, and we refer to [44] or [85] for the details.

We remark that we recover the zero-bias transform letting $C \to 0$. Moreover, the translation of (4.33) in terms of a fixed point equation in law is

$$X \sim \mathscr{H}(\Phi_C, \gamma) \iff X \stackrel{\mathcal{L}}{=} X^{(C)}$$

Unfortunately, due to the non-linearity of ρ_C, the application to sums of i.i.d. random variables fails : the C-bias transform of S_n is not immediate to find, and the replacement at random of one term of the sum by an independent C-biased term does not give the result.

4.8 Appendix : Stein's estimates

We develop here the equivalent of the Stein's estimates that are relevant in our case by carefully adapting the steps of Stein [94].

4.8.1 Basic estimates

Lemma 4.8.1. *For $q \geqslant 0$, define the random variable Z_q and the polynomial P_q by*

$$\begin{aligned} P_q(x) &:= \frac{x^2}{2} + q\frac{x^4}{4} \\ \mathbb{P}(Z_q \leqslant x) &:= \int_{-\infty}^{x} e^{-P_q(y)} \frac{dy}{z_q} \quad \textit{with} \quad z_q := \int_{\mathbb{R}} e^{-P_q(y)} dy \end{aligned}$$

Then,

$$\forall\, x > 0, \quad \mathbb{P}(Z_q \geqslant x) \leqslant \frac{e^{-P_q(x)}}{z_q P_q'(x)} \tag{4.34}$$

$$\forall\, x < 0, \quad \mathbb{P}(Z_q \leqslant x) \leqslant \frac{e^{-P_q(x)}}{z_q P_q'(|x|)} \tag{4.35}$$

Proof. As $P_q''(x) = 1 + 3qx^2 > 0$, the function $P_q' : x \mapsto x + qx^3$ is strictly increasing on $\mathbb{R}$ and we can write for $x > 0$

$$z_q \mathbb{P}\left(Z_q \geqslant x\right) = \int_x^{+\infty} e^{-P_q(y)} dy < \int_x^{+\infty} \frac{P_q'(u)}{P_q'(x)} e^{-P_q(y)} dy = \frac{e^{-P_q(x)}}{P_q'(x)}$$

and analogously, as $Z_q \stackrel{\mathcal{L}}{=} -Z_q$, for $x < 0$

$$\mathbb{P}\left(Z_q \leqslant x\right) \leqslant \frac{e^{-P_q(x)}}{z_q P_q'(|x|)}$$

□

Corollary 4.8.2. *Set*

$$\rho_\gamma(x) := \frac{1}{\gamma} P_{q_\gamma}'\left(\frac{x}{\gamma}\right) = \frac{1}{\gamma^2} x + \frac{C}{\gamma^8} x^3 =: ax + bx^3 \tag{4.36}$$

Then,

$$\forall\, x > 0, \qquad \mathbb{P}\left(H_\gamma \geqslant x\right) \leqslant \frac{f_{H_\gamma}(x)}{\rho_\gamma(x)} \tag{4.37}$$

$$\forall\, x < 0, \qquad \mathbb{P}\left(H_\gamma \leqslant x\right) \leqslant \frac{f_{H_\gamma}(x)}{\rho_\gamma(|x|)} \tag{4.38}$$

Proof. It is clear that

$$H_\gamma \stackrel{\mathcal{L}}{=} \gamma Z_{q_\gamma} \quad \text{with} \quad q_\gamma := \frac{C}{\gamma^4}$$

Moreover, with c_γ defined in (3.20),

$$\begin{aligned} z_{q_\gamma} &:= \sqrt{2\pi}\mathbb{E}\left(e^{-q_\gamma \frac{G^4}{4}}\right) = \sqrt{2\pi}\mathbb{E}\left(e^{-\frac{C}{4}\left(\frac{G}{\gamma}\right)^4}\right) = \frac{c_\gamma}{\gamma} \\ f_{H_\gamma}(x) &:= \frac{1}{c_\gamma} e^{-\frac{C}{4}\left(\frac{x}{\gamma^2}\right)^4} e^{-\frac{1}{2}\left(\frac{x}{\gamma}\right)^2} = \frac{1}{\gamma z_{q_\gamma}} e^{-P_{q_\gamma}\left(\frac{x}{\gamma}\right)} \end{aligned}$$

Hence, for $x > 0$

$$\mathbb{P}\left(H_\gamma \geqslant x\right) = \mathbb{P}\left(Z_{q_\gamma} \geqslant \frac{x}{\gamma}\right) \leqslant \frac{e^{-P_{q_\gamma}\left(\frac{x}{\gamma}\right)}}{z_{q_\gamma} P_{q_\gamma}'\left(\frac{x}{\gamma}\right)} = \frac{f_{H_\gamma}(x)}{\frac{1}{\gamma} P_{q_\gamma}'\left(\frac{x}{\gamma}\right)}$$

□

Lemma 4.8.3. *We have for all $x \in \mathbb{R}$*

$$\mathbb{E}\left(H_\gamma \mathbb{1}_{\{H_\gamma \geqslant x\}}\right) \leqslant \frac{x f_{H_\gamma}(x)}{\rho_\gamma(x)} \tag{4.39}$$

Proof. The fact that this last inequality is symmetric compared for example to (4.38) comes from the fact that $R(x) := \rho_\gamma(x)/x = a + bx^2 = R(|x|)$ and the fact that the function on the left hand side is odd.

As $(f_{H_\gamma}/R)' = -f_{H_\gamma}\left(R'/R^2 + \rho_\gamma/R\right)$, and $\lim_{+\infty} f_{H_\gamma}/R = 0$, we have

$$\frac{x f_{H_\gamma}(x)}{\rho_\gamma(x)} = \int_x^{+\infty} \left(R'/R^2 + \rho_\gamma/R\right) f_{H_\gamma} = \mathbb{E}\left(\left(R'/R^2 + \rho_\gamma/R\right)(H_\gamma)\mathbb{1}_{\{H_\gamma \geqslant x\}}\right)$$

But $R(x) := \rho_\gamma(x)/x$, so $\rho_\gamma(x)/R(x) = x$ and it remains to show that

$$\mathbb{E}\left(\frac{R'(H_\gamma)}{R(H_\gamma)^2}\mathbb{1}_{\{H_\gamma \geqslant x\}}\right) \geqslant 0$$

which is clearly the case for $x \geqslant 0$ since $R'(x) = 2bx$.

For $x \leqslant 0$, set $r := R'/R^2$ and remark that $r(-x) = -r(x)$. As $H_\gamma \stackrel{\mathcal{L}}{=} -H_\gamma$, this implies that $r(H_\gamma) \stackrel{\mathcal{L}}{=} -r(H_\gamma)$ and in particular that $\mathbb{E}\left(r(H_\gamma)\right) = 0$. Then

$$\begin{aligned}
\mathbb{E}\left(r(H_\gamma)\mathbb{1}_{\{H_\gamma \geqslant x\}}\right) &= \mathbb{E}\left(r(H_\gamma)\mathbb{1}_{\{H_\gamma \geqslant -|x|\}}\right) = -\mathbb{E}\left(r(-H_\gamma)\mathbb{1}_{\{-H_\gamma \leqslant |x|\}}\right) \\
&= -\mathbb{E}\left(r(H_\gamma)\mathbb{1}_{\{H_\gamma \leqslant |x|\}}\right) \quad \text{as} \quad r(H_\gamma) \stackrel{\mathcal{L}}{=} -r(H_\gamma) \\
&= \mathbb{E}\left(r(H_\gamma)\mathbb{1}_{\{H_\gamma \geqslant |x|\}}\right) \quad \text{as} \quad \mathbb{E}\left(r(H_\gamma)\right) = 0 \\
&\geqslant r(|x|)\mathbb{P}\left(H_\gamma \geqslant |x|\right) \geqslant 0
\end{aligned}$$

□

Replacing the function r by $x \mapsto x$ in the last equalities gives $\mathbb{E}\left(H_\gamma \mathbb{1}_{\{H_\gamma \geqslant x\}}\right) \geqslant 0$ for all $x \in \mathbb{R}$.

Lemma 4.8.4. *We have for all $x \in \mathbb{R}$*

$$\mathbb{P}\left(H_\gamma \leqslant x\right) \geqslant -\frac{\rho_\gamma(x)}{\rho'_\gamma(x) + \rho_\gamma(x)^2} f_{H_\gamma}(x) \tag{4.40}$$

$$\mathbb{P}\left(H_\gamma \geqslant x\right) \geqslant \frac{\rho_\gamma(x)}{\rho'_\gamma(x) + \rho_\gamma(x)^2} f_{H_\gamma}(x) \tag{4.41}$$

Proof. We first remark that

$$\begin{aligned}
\mathbb{P}\left(H_\gamma \leqslant x\right) + \frac{\rho_\gamma(x)}{\rho'_\gamma(x) + \rho_\gamma(x)^2} f_{H_\gamma}(x) &= \int_{-\infty}^{x} \left(f_{H_\gamma}(u) + \frac{d}{du}\left(\frac{\rho_\gamma f_{H_\gamma}}{\rho'_\gamma + \rho_\gamma^2}\right)(u)\right) du \\
\mathbb{P}\left(H_\gamma \geqslant x\right) - \frac{\rho_\gamma(x)}{\rho'_\gamma(x) + \rho_\gamma(x)^2} f_{H_\gamma}(x) &= \int_{x}^{+\infty} \left(f_{H_\gamma}(u) + \frac{d}{du}\left(\frac{\rho_\gamma f_{H_\gamma}}{\rho'_\gamma + \rho_\gamma^2}\right)(u)\right) du
\end{aligned}$$

Hence, it is sufficient to prove that

$$1 + \frac{1}{f_{H_\gamma}(x)}\frac{d}{dx}\left(\frac{\rho_\gamma f_{H_\gamma}}{\rho'_\gamma + \rho_\gamma^2}\right)(x) \geqslant 0$$

But

$$\begin{aligned}1+\frac{1}{f_{H_\gamma}}\left(\frac{\rho_\gamma f_{H_\gamma}}{\rho'_\gamma+\rho_\gamma^2}\right)' &= 1+\left(\frac{\rho_\gamma}{\rho'_\gamma+\rho_\gamma^2}\right)'-\rho_\gamma\left(\frac{\rho_\gamma}{\rho'_\gamma+\rho_\gamma^2}\right)=\left(\frac{\rho_\gamma}{\rho'_\gamma+\rho_\gamma^2}\right)'+\frac{\rho'_\gamma}{\rho'_\gamma+\rho_\gamma^2}\\ &= \frac{1}{\rho_\gamma}\left(\frac{\rho_\gamma^2}{\rho'_\gamma+\rho_\gamma^2}\right)'=\frac{1}{\rho_\gamma}\left(\frac{1}{1-(1/\rho_\gamma)'}\right)'\\ &= \frac{1}{\rho_\gamma}\left(\frac{1}{\rho_\gamma}\right)''\frac{1}{\left(1-(1/\rho_\gamma)'\right)^2}=\frac{2(\rho'_\gamma)^2-\rho_\gamma\rho''_\gamma}{\rho_\gamma^4\left(1-(1/\rho_\gamma)'\right)^2}\end{aligned}$$

It is now sufficient to prove that $2(\rho'_\gamma)^2-\rho_\gamma\rho''_\gamma\geqslant 0$. Setting $X:=x^2$, we have

$$\begin{aligned}2(\rho'_\gamma)^2(x)-\rho_\gamma(x)\rho''_\gamma(x) &= 2(a+3bx^2)^2-6bx(ax+bx^3)=2(a+3bX)^2-6(abX+b^2X^2)\\ &= 12(bX)^2+6abX+2a^2=2a^2\left(6\left(\frac{bX}{a}\right)^2+3\left(\frac{bX}{a}\right)+1\right)\end{aligned}$$

As the function $t\mapsto 6t^2+3t+1$ is positive on $\mathbb{R}$, we finally have the result.

□

4.8.2 Operator norms estimates

Lemma 4.8.5 (Operator norms estimates)**.** *For $h\in\mathcal{H}_\Phi$, let $h_{H_n}:=h-\mathbb{E}\left(h(H_n)\right)$ and let $g:=\mathcal{L}_{H_n}^{-1}h_{H_n}$ be the solution of the Stein's equation that goes to 0 in $\pm\infty$, i.e.*

$$\mathcal{L}_{H_n}g=h_{H_n} \tag{4.42}$$

with

$$\mathcal{L}_{H_n}g(x):=g'(x)-\left(\frac{x}{\gamma_n^2}+C\frac{x^3}{\gamma_n^8}\right)g(x)=:g'(x)-\rho_{\gamma_n}(x)g(x) \tag{4.43}$$

Then :

1. *If h is bounded ($\|h\|_\infty<\infty$), setting $c_1:=\sqrt{\frac{\pi}{2}}\mathbb{E}\left(\exp\left(-\frac{CG^4}{4}\right)\right)$ and D for the operator of differentiation,*

$$\left\|\mathcal{L}_{H_n}^{-1}h_{H_n}\right\|_\infty\leqslant c_1\gamma_n\left\|h_{H_n}\right\|_\infty \tag{4.44}$$

$$\left\|D\mathcal{L}_{H_n}^{-1}h_{H_n}\right\|_\infty\leqslant 2\left\|h_{H_n}\right\|_\infty \tag{4.45}$$

2. *If h is absolutely continuous ($\|Dh\|_\infty<\infty$) and $\int_\mathbb{R}|h'|<\infty$,*

$$\left\|D^2\mathcal{L}_{H_n}^{-1}h_{H_n}\right\|_\infty\leqslant 4\left\|Dh_{H_n}\right\|_\infty \tag{4.46}$$

Proof. 1. • Using the representation (4.23) and for $x > 0$, we have

$$\begin{aligned}
\left|\mathcal{L}_{H_n}^{-1} h_{H_n}(x)\right| &\leqslant \frac{\mathbb{E}\left(|h_{H_n}(H_n)|\,\mathbb{1}_{\{H_n \geqslant x\}}\right)}{f_{H_n}(x)} \\
&\leqslant \left\|h_{H_n}\right\|_\infty \frac{\mathbb{E}\left(\mathbb{1}_{\{H_n \geqslant x\}}\right)}{f_{H_n}(x)} \\
&\leqslant \left\|h_{H_n}\right\|_\infty \sup_{x>0} \frac{\mathbb{P}\left(H_n \geqslant x\right)}{f_{H_n}(x)} \\
&= \frac{\mathbb{P}\left(H_n \geqslant 0\right)}{f_{H_n}(0)} \left\|h_{H_n}\right\|_\infty = \frac{1/2}{1/(z_{q_{\gamma_n}}\gamma_n)} \left\|h_{H_n}\right\|_\infty
\end{aligned}$$

As $H_\gamma \stackrel{\mathcal{L}}{=} -H_\gamma$, we have $\mathbb{P}\left(H_\gamma \geqslant 0\right) = 1/2$. The fact that $x \mapsto \mathbb{P}\left(H_n \leqslant x\right)/f_{H_n}(x)$ is maximum in $x = 0$ follows from (4.38) : if, for $H_\gamma \sim \mathscr{H}(\Phi_C, \gamma)$ we form the functions

$$\begin{aligned}
s_\gamma^{(+)} &: x > 0 \mapsto \frac{\mathbb{P}\left(H_\gamma \geqslant x\right)}{f_{H_\gamma}(x)} \\
s_\gamma^{(-)} &: x < 0 \mapsto \frac{\mathbb{P}\left(H_\gamma \leqslant x\right)}{f_{H_\gamma}(x)}
\end{aligned}$$

we get for $\varepsilon \in \{\pm 1\}$

$$\frac{ds_\gamma^{(\varepsilon)}}{dx}(x) = -\varepsilon + \frac{1}{\gamma} P'_{q_\gamma}\left(\frac{x}{\gamma}\right) s_\gamma^{(\varepsilon)}(x)$$

Hence, by (4.37) and (4.38), $s_\gamma^{(+)}$ is decreasing on $\mathbb{R}_+$ (and $s_\gamma^{(-)}$ is increasing on $\mathbb{R}_-$), so that they attain their maxima in 0. Thus, for $x > 0$

$$\left|\mathcal{L}_{H_n}^{-1} h_{H_n}(x)\right| \leqslant \frac{z_{q_{\gamma_n}}\gamma_n}{2} \left\|h_{H_n}\right\|_\infty = \frac{\gamma_n\sqrt{2\pi}}{2} \mathbb{E}\left(\exp\left(-\frac{C}{4}\frac{G^4}{\gamma_n^4}\right)\right) \left\|h_{H_n}\right\|_\infty$$

It is clear that $\gamma \in \mathbb{R}_+ \mapsto \mathbb{E}\left(\exp\left(-\frac{C}{4}\frac{G^4}{\gamma^4}\right)\right)$ is decreasing, so that

$$\max_{n\geqslant 1}\left\{\sqrt{\frac{\pi}{2}}\mathbb{E}\left(\exp\left(-\frac{C}{4}\frac{G^4}{\gamma_n^4}\right)\right)\right\} = \sqrt{\frac{\pi}{2}}\mathbb{E}\left(\exp\left(-\frac{CG^4}{4}\right)\right) =: c_1$$

For $x < 0$, the same exact arguments apply using the other representation of $\mathcal{L}_{H_n}^{-1} h_{H_n}$ and $s_\gamma^{(-)}$, leading to

$$\left\|\mathcal{L}_{H_n}^{-1} h_{H_n}\right\|_\infty \leqslant c_1 \gamma_n \left\|h_{H_n}\right\|_\infty$$

- As $g = \mathcal{L}_{H_n}^{-1} h_{H_\gamma}$, is the solution of the Stein's equation (4.42), we have

$$D\mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma}(x) = h_{H_\gamma}(x) + \frac{1}{\gamma} P'_{q_\gamma}\left(\frac{x}{\gamma}\right) \mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma}(x)$$

  Thus, for $x \geqslant 0$

$$\begin{aligned}\left|D\mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma}(x)\right| &\leqslant \left|h_{H_\gamma}(x)\right| + \left|\frac{1}{\gamma} P'_{q_\gamma}\left(\frac{x}{\gamma}\right)\right| \left|\mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma}(x)\right| \\ &\leqslant \left\|h_{H_\gamma}\right\|_\infty \left(1 + \sup_{x>0}\left\{\left|\frac{1}{\gamma} P'_{q_\gamma}\left(\frac{x}{\gamma}\right)\right| \frac{\mathbb{P}\left(H_\gamma \geqslant x\right)}{f_{H_\gamma}(x)}\right\}\right) \\ &\leqslant 2\left\|h_{H_\gamma}\right\|_\infty \quad \text{by (4.37)}\end{aligned}$$

  We proceed the same for $x < 0$.

2. To prove that $\left\|D^2\mathcal{L}_{H_n}^{-1} h_{H_n}\right\|_\infty \leqslant 4\left\|Dh_{H_n}\right\|_\infty$, we express $D^2\mathcal{L}_{H_n}^{-1} h_{H_n}$ in terms of h'. First, we differentiate (4.42),

$$\begin{aligned}D^2\mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma} &= D\left(\rho_\gamma \mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma} + h_{H_\gamma}\right) = \rho'_\gamma \mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma} + \rho_\gamma D\mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma} + h' \\ &= \left(\rho'_\gamma + \rho_\gamma^2\right) \mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma} + \rho_\gamma h_{H_\gamma} + h' \end{aligned} \tag{4.47}$$

We already have

$$\begin{aligned}h_{H_\gamma}(x) &= \mathbb{E}\left(h(x) - h(H_\gamma)\right) \\ &= \mathbb{E}\left(\int_{H_\gamma}^x h'(u)du\, \mathbb{1}_{\{H_\gamma < x\}}\right) + \mathbb{E}\left(\int_{H_\gamma}^x h'(u)du\, \mathbb{1}_{\{H_\gamma > x\}}\right) \\ &= \mathbb{E}\left(\int_{\mathbb{R}} h'(u)\left(\mathbb{1}_{\{H_\gamma < u < x\}} - \mathbb{1}_{\{H_\gamma > u > x\}}\right) du\right) \\ &= \int_{\mathbb{R}} h'(u)\mathbb{E}\left(\mathbb{1}_{\{H_\gamma < u < x\}} - \mathbb{1}_{\{H_\gamma > u > x\}}\right) du =: \int_{\mathbb{R}} h'(u) K_H(x,u) du\end{aligned}$$

where the last equality comes from the Fubini theorem (since $\int_{\mathbb{R}} |h'| < \infty$), and where

$$K_H(x,u) := \mathbb{E}\left(\mathbb{1}_{\{H_\gamma < u < x\}} - \mathbb{1}_{\{H_\gamma > u > x\}}\right) := \mathbb{P}\left(H_\gamma < u\right)\mathbb{1}_{\{u<x\}} - \mathbb{P}\left(H_\gamma > u\right)\mathbb{1}_{\{u>x\}} \tag{4.48}$$

We moreover have

$$\begin{aligned}\mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma}(x) &= \frac{1}{f_{H_\gamma}(x)} \mathbb{E}\left(h_{H_\gamma}(H_\gamma)\mathbb{1}_{\{H_\gamma \leqslant x\}}\right) = -\frac{1}{f_{H_\gamma}(x)} \mathbb{E}\left(h_{H_\gamma}(H_\gamma)\mathbb{1}_{\{H_\gamma \geqslant x\}}\right) \\ &= \frac{1}{2f_{H_\gamma}(x)} \mathbb{E}\left(h_{H_\gamma}(H_\gamma)\left[\mathbb{1}_{\{H_\gamma \leqslant x\}} - \mathbb{1}_{\{H_\gamma \geqslant x\}}\right]\right) \\ &=: \frac{1}{f_{H_\gamma}(x)} \mathbb{E}\left(I\left(x, H_\gamma\right) h_{H_\gamma}(H_\gamma)\right)\end{aligned}$$

where

$$I(x,y) := \frac{1}{2}\left(\mathbb{1}_{\{y<x\}} - \mathbb{1}_{\{y>x\}}\right)$$

It follows that

$$\begin{aligned}
\mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma}(x) &= \frac{1}{f_{H_\gamma}(x)} \mathbb{E}\Big(I\left(x, H_\gamma\right) h_{H_\gamma}(H_\gamma)\Big) \\
&= \frac{1}{f_{H_\gamma}(x)} \mathbb{E}\left(I\left(x, H_\gamma\right) \int_{\mathbb{R}} K_H(H_\gamma, u) h'(u) du\right) \\
&= \int_{\mathbb{R}} \frac{\mathbb{E}\left(I\left(x, H_\gamma\right) K_H(H_\gamma, u)\right)}{f_{H_\gamma}(x)} h'(u)\, du =: \int_{\mathbb{R}} \widetilde{K}_H(x,u) h'(u)\, du
\end{aligned}$$

where the last equality comes from the Fubini theorem.

Let $H_\gamma^{(1)}, H_\gamma^{(2)}$ be two independent random variables equal in law to H_γ. Then,

$$\begin{aligned}
& f_{H_\gamma}(x)\, \widetilde{K}_H(x,u) := \mathbb{E}\left(I\left(x, H_\gamma\right) K_H(H_\gamma, u)\right) = \mathbb{E}\left(I\left(x, H_\gamma\right) K_H(H_\gamma, u) \left[\mathbb{1}_{\{x<u\}} + \mathbb{1}_{\{u<x\}}\right]\right) \\
&= \frac{1}{2}\mathbb{E}\left(\left[\mathbb{1}_{\left\{H_\gamma^{(1)} \leqslant x\right\}} - \mathbb{1}_{\left\{H_\gamma^{(1)} \geqslant x\right\}}\right]\left[\mathbb{1}_{\left\{H_\gamma^{(2)} < u < H_\gamma^{(1)}\right\}} - \mathbb{1}_{\left\{H_\gamma^{(2)} > u > H_\gamma^{(1)}\right\}}\right]\left[\mathbb{1}_{\{x<u\}} + \mathbb{1}_{\{u<x\}}\right]\right) \\
&= \frac{1}{2}\mathbb{E}\Bigg(0 + \mathbb{1}_{\left\{H_\gamma^{(2)} > u > H_\gamma^{(1)} > x\right\}} - \mathbb{1}_{\left\{H_\gamma^{(2)} < u < H_\gamma^{(1)},\, H_\gamma^{(1)} > x,\, x<u\right\}} - \mathbb{1}_{\left\{H_\gamma^{(2)} > u > x > H_\gamma^{(1)}\right\}} \\
&\qquad + \mathbb{1}_{\left\{H_\gamma^{(2)} < u < H_\gamma^{(1)} < x\right\}} + 0 - \mathbb{1}_{\left\{H_\gamma^{(2)} < u < x < H_\gamma^{(1)}\right\}} - \mathbb{1}_{\left\{H_\gamma^{(2)} > u > H_\gamma^{(1)},\, H_\gamma^{(1)} < x,\, x>u\right\}}\Bigg) \\
&= \frac{1}{2}\mathbb{E}\Bigg(\mathbb{1}_{\left\{H_\gamma^{(2)} > u > H_\gamma^{(1)} > x\right\}} - \mathbb{1}_{\left\{H_\gamma^{(2)} < x < u < H_\gamma^{(1)}\right\}} - \mathbb{1}_{\left\{x < H_\gamma^{(2)} < u < H_\gamma^{(1)}\right\}} - \mathbb{1}_{\left\{H_\gamma^{(2)} > u > x > H_\gamma^{(1)}\right\}} \\
&\qquad + \mathbb{1}_{\left\{H_\gamma^{(2)} < u < H_\gamma^{(1)} < x\right\}} - \mathbb{1}_{\left\{H_\gamma^{(2)} < u < x < H_\gamma^{(1)}\right\}} - \mathbb{1}_{\left\{H_\gamma^{(2)} > x > u > H_\gamma^{(1)}\right\}} - \mathbb{1}_{\left\{x > H_\gamma^{(2)} > u > H_\gamma^{(1)}\right\}}\Bigg) \\
&= -\mathbb{E}\left(\mathbb{1}_{\left\{H_\gamma^{(2)} < u < x < H_\gamma^{(1)}\right\}} + \mathbb{1}_{\left\{H_\gamma^{(2)} > u > x > H_\gamma^{(1)}\right\}}\right)
\end{aligned}$$

this last equality coming from the exchangeability of $(H_\gamma^{(1)}, H_\gamma^{(2)})$.

Finally, we get

$$\tilde{K}_H(x,u) = -\frac{1}{f_{H_\gamma}(x)} \mathbb{E}\left(\mathbb{1}_{\left\{H_\gamma^{(2)} < u < x < H_\gamma^{(1)}\right\}} + \mathbb{1}_{\left\{H_\gamma^{(2)} > u > x > H_\gamma^{(1)}\right\}}\right) \tag{4.49}$$

From (4.47), (4.48) and (4.49), and setting

$$K \star h(x) := \int_{\mathbb{R}} K(x,y) h(y) dy$$

we get

$$\begin{aligned}
D^2 \mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma} &= h' + \left(\rho_\gamma' + \rho_\gamma^2\right) \mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma} + \rho_\gamma h_{H_\gamma} = h' + \left(\rho_\gamma' + \rho_\gamma^2\right) \tilde{K}_H \star h' + \rho_\gamma K_H \star h' \\
&=: h' + K \star h' \text{ with } \; K(x,y) := \left(\rho_\gamma' + \rho_\gamma^2\right)(x) \tilde{K}_H(x,y) + \rho_\gamma(x) K_H(x,y)
\end{aligned}$$

Setting $F_{H_\gamma}(x) := \mathbb{P}(H_\gamma \leqslant x)$ and $\overline{F}_{H_\gamma}(x) := \mathbb{P}(H_\gamma \geqslant x)$, this last operator writes

$$\begin{aligned}
K \star h'(x) &:= \int_{\mathbb{R}} \Big(\left(\rho_\gamma'(x) + \rho_\gamma^2(x)\right) \tilde{K}_H(x,y) + \rho_\gamma(x) K_H(x,y) \Big) h'(y)dy \\
&= \left(\rho_\gamma - \frac{\rho_\gamma' + \rho_\gamma^2}{f_{H_\gamma}} \overline{F}_{H_\gamma} \right)(x) \int_{-\infty}^{x} F_{H_\gamma}(y) h'(y) dy \\
&\quad - \left(\rho_\gamma + \frac{\rho_\gamma' + \rho_\gamma^2}{f_{H_\gamma}} F_{H_\gamma} \right)(x) \int_{x}^{+\infty} \overline{F}_{H_\gamma}(y) h'(y) dy \\
&=: -K^{(-)} \star h'(x) + K^{(+)} \star h'(x)
\end{aligned}$$

Using (4.40) and the obvious fact that F_{H_γ} and $\overline{F}_{H_\gamma}$ are positive, we have

$$\left|K \star h'(x)\right| \leqslant K^{(-)} \star \left|h'\right|(x) + K^{(+)} \star \left|h'\right|(x) \leqslant \left\|h'\right\|_\infty \left(K^{(-)} + K^{(+)}\right) \star \mathbb{1}(x)$$

with

$$\begin{aligned}
\left(K^{(-)} + K^{(+)}\right) \star \mathbb{1}(x) &= -\left(\rho_\gamma - \frac{\rho_\gamma' + \rho_\gamma^2}{f_{H_\gamma}} \overline{F}_{H_\gamma} \right)(x) \int_{-\infty}^{x} F_{H_\gamma}(y) dy \\
&\quad + \left(\rho_\gamma + \frac{\rho_\gamma' + \rho_\gamma^2}{f_{H_\gamma}} F_{H_\gamma} \right)(x) \int_{x}^{+\infty} \overline{F}_{H_\gamma}(y) dy
\end{aligned}$$

Using the Fubini theorem and $\mathbb{E}(H_\gamma) = 0$, we have

$$\int_{-\infty}^{x} F_{H_\gamma}(u) du = \int_{\mathbb{R}} \mathbb{E}\left(\mathbb{1}_{\{H_\gamma \leqslant u \leqslant x\}}\right) du = \mathbb{E}\left((x - H_\gamma)_+\right) = x F_{H_\gamma}(x) + \mathbb{E}\left(H_\gamma \mathbb{1}_{\{H_\gamma \geqslant x\}}\right)$$

In addition,

$$\begin{aligned}
\int_{-\infty}^{x} F_{H_\gamma} - \int_{x}^{+\infty} \overline{F}_{H_\gamma} &= \int_{\mathbb{R}} \mathbb{E}\left(\mathbb{1}_{\{H_\gamma \leqslant u \leqslant x\}} - \mathbb{1}_{\{H_\gamma \geqslant u \geqslant x\}}\right) du \\
&= \mathbb{E}\left((x - H_\gamma)_+ - (H_\gamma - x)_+\right) = x
\end{aligned}$$

We thus deduce

$$\begin{aligned}
\left(K^{(-)} + K^{(+)}\right) \star \mathbb{1}(x) &= \frac{\rho_\gamma'(x) + \rho_\gamma^2(x)}{f_{H_\gamma}(x)} \mathbb{E}\left(H_\gamma \mathbb{1}_{\{H_\gamma \geqslant x\}}\right) - x\rho_\gamma(x) \\
&\leqslant \frac{\rho_\gamma'(x) + \rho_\gamma^2(x)}{\rho_\gamma(x)/x} - x\rho_\gamma(x) \quad \text{using (4.39)} \\
&= \frac{x\rho_\gamma'(x)}{\rho_\gamma(x)} = 3 - \frac{1}{1 + Cx^2/\gamma^6} \\
&\leqslant 3
\end{aligned}$$

Finally, we get

$$\begin{aligned}
\left|D^2 \mathcal{L}_{H_\gamma}^{-1} h_{H_\gamma}(x)\right| &= \left|h'(x) + K \star h'(x)\right| \leqslant \left|h'(x)\right| + \left|K \star h'(x)\right| \\
&\leqslant \left\|h'\right\|_\infty \left(1 + \left(K^{(-)} + K^{(+)}\right) \star \mathbb{1}(x)\right) \\
&\leqslant 4 \left\|h'\right\|_\infty
\end{aligned}$$

□

Chapter 5

Conclusion

The Keating-Snaith philosophy was a revolutionary approach to think Number Theory from a conjectural point of view. Its counterpart that consists in starting from Number Theory to go to random matrices is also an important aspect of the correspondance.

The fact that the moments conjecture was stated in a non probabilistic language despite its clear probabilistic framework was a difficulty that led Jacod, Kowalski and Nikeghbali to initiate the systematic study of sequences of random variables converging in such a sense, with the goal to determine which results are the consequence of a general phenomenon and what is particular to a specific example.

From this point of view, mod-* convergence is far from having revealed all its mysteries. The probabilistic interpretation in terms of distance to a "canonical" distribution constructed by means of a penalisation with the limiting function Φ can be of a certain help in understanding which type of convergence we are dealing with and to compare it with the convergence in distribution, but also for more applied purposes since such a second-order distribution can be used to make approximation in distribution. From the point of view of the technique, being able to use purely probabilistic methods such as Stein's can also be useful.

The philosophy of additional randomisation explains the form that takes the limiting functions in presence of a natural independent model that can be achieved by such an operation. In particular, as mod-* convergence acts as a correction to a dependence that vanishes at the first order with a renormalisation, this is natural to look for an operation that introduces the independance (or, at least, an additional property that is not present in the initial sequence), to better understand how to get rid of it. Looking for such a randomized model in the case of the moments conjecture is now natural, and understanding if every model can be thought of in this way is one of its ambitious prolongations.

Last, constructing models that share the same mod-* fluctuations as a certain sequence of random variables is also an important aspect of mod-* convergence. The understanding of these models can give interesting informations on the behaviour of the random variables, as this was the case for $\omega(U_n)$. Extending such a construction to other random variables could reveal unexpected links with other fields of mathematics a priori unrelated to it.

Bibliography

[1] G. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices,* Cambridge University press, (2009).

[2] R. Arratia, A. D. Barbour, and S. Tavaré, *Logarithmic combinatorial structures, a probabilistic approach,* EMS Monographs in Mathematics, Zürich, Europ. Math. Soc. (2003).

[3] A. D. Barbour, *Asymptotic expansions based on smooth functions in the Central Limit Theorem,* Proba. Th. and Rel. Fields, **72** (1986), p 289-303.

[4] A. D. Barbour, E. Kowalski, A. Nikeghbali, *Mod-discrete expansions,* preprint (2009), available at `http://arxiv.org/pdf/0912.1886`

[5] Y. Barhoumi-Andréani, *Stein's method and mod-* convergence,* preprint (**2013**).

[6] Y. Barhoumi-Andréani, *Mod-* convergence with change of probability : reproducing the fluctuations in the Sathé-Selberg theorem,* preprint (**2013**).

[7] Y. Barhoumi-Andréani, C. P. Hughes, J. Najnudel, A. Nikeghbali, *On the number of zeros of linear combinations of independent characteristic polynomials of random unitary matrices,* preprint `https://www.math.uzh.ch/fileadmin/user/ashkan/publikation/cw.pdf` (**2013**).

[8] E. W. Barnes, *The theory of the G-function,* Quarterly Journ. Pure and Appl. Math. **vol 31** (1900), pp. 264-314.

[9] J. Ben Hough, M. Krishnapur, Y. Peres, B. Virag, *Zeros of Gaussian Analytic Functions and Determinantal Point Processes,* American Mathematical Society, Providence (2009).

[10] E. Bogomolny, O. Bohigas, P. Leboeuf *Quantum chaotic dynamics and random polynomials,* Journal of Statistical Physics, **85**, Nos. 5/6 (1996), p 639–679.

[11] E. Bombieri, D.A. Hejhal, *On the distribution of zeros of linear combinations of Euler products,* Duke Math. J., **80** (1995), p 821–862.

[12] A. Borodin, G. Olshanski, *Infinite random matrices and ergodic measures,* Commun. Math. Phys., **223** (2001), p 87-123.

[13] P. Bourgade, *A propos des matrices aléatoires et des fonctions L,* Ph.D. Thesis (2009), `http://tel.archives-ouvertes.fr/tel-00373735/fr`.

[14] P. Bourgade, J. Najnudel , A. Nikeghbali, *A unitary extension of virtual permutations,* To appear in IMRN.

[15] P. Bourgade, A. Nikeghbali, A. Rouault, *Ewens measures on compact groups and hypergeometric kernels,* Séminaire de Probabilités, **Vol. 43**, pp. 351-377, LNM 2006 (2011).

[16] G. Ben Arous, K. Dang, *On fluctuations of eigenvalues of random permutation matrices,* preprint `http://arXiv:1106.2108`.

[17] A. C. Berry, *The accuracy of the Gaussian approximation to the sum of independent variates,* Transactions of the American Mathematical Society **vol. 49** (1), pp. 122-136 (1941).

[18] P. Biane, J. Pitman, M. Yor, *Probability laws related to the Jacobi theta and Riemann zeta function and Brownian excursions,* Bull. Amer. Math. Soc. **vol. 38** (2001), pp. 435-465.

[19] P. Bourgade, C. Hughes, A. Nikeghbali, M. Yor, *The characteristic polynomial of a random unitary matrix : a probabilistic approach,* Duke Math. J., **145** (2008), p 45–69.

[20] P. Bourgade, A. Nikeghbali, A. Rouault, *Ewens measures on compact groups and hypergeometric kernels,* Séminaire de Probabilités, **XLIII**, Lecture Notes in Math. **2006** (2010), pp. 351-377.

[21] M. Bramson, O. Zeitouni, *Tightness of the recentered maximum of the two-dimensional discrete Gaussian Free Field,* preprint available at `http://arxiv.org/abs/1009.3443`.

[22] D. Bump, A. Gamburd, *On the averages of characteristic polynomials from classical groups,* Commun. Math. Phys., **vol. 265** (2006), pp. 227-274.

[23] L.H.Y. Chen, *Poisson approximation for dependent trials,* Annals of Probability **vol. 3** (3), pp. 534-545 (1975).

[24] R. Chhaibi, *Modèle de Littelmann pour cristaux géométriques, fonctions de Whittaker sur des groupes de Lie et mouvement brownien,* Ph.D. thesis, (2012).

[25] J. B. Conrey, D. W. Farmer, J.P. Keating, M. O. Rubinstein, N.C. Snaith, *Autocorrelation of random matrix polynomials,* Commun. Math. Phys. **vol. 237** nb 3 (2003), pp. 365-395

[26] J. B. Conrey, P. J. Forrester, N.C. Snaith, *Averages of ratios of characteristic polynomials for the compact classical groups,* International Mathematics Research Notices, **vol. 7** (2005), pp. 397-431.

[27] J. B. Conrey, M. Gonek, *High moments of the Riemann zeta-function,* Duke Math. J. **Vol. 107**, Number 3 (2001), pp. 577-604.

[28] J. B. Conrey, A. Ghosh, *A conjecture for the sixth power moment of the Riemann zeta-function,* Int. Math. Res. Not. **Vol. 15** (1998), pp. 775-780.

[29] F. Delbaen, E. Kowalski, A. Nikeghbali, *Mod-ϕ convergence,* preprint (2011), available at `http://arxiv.org/abs/1107.5657`

[30] P. Diaconis, *Patterns in eigenvalues : the 70th Josiah Willard Gibbs lecture,* Bulletin of the AMS (new series), **vol. 40**, nb 2 (2003), pp. 155-178.

[31] P. Diaconis, A. Gamburd, *Random matrices, magic squares and matching polynomials,* Electronic Journal of combinatorics, **Vol. 11**, Part 2 (2004), # R2.

[32] P. Diaconis, M. Shahshahani, *The subgroup algorithm for generating uniform random variables,* Prob. Eng. Inf. Sc., **vol. 1** (1987), pp. 15-32.

[33] P. Diaconis, M. Shahshahani, *On the Eigenvalues of Random Matrices,* Journal of Applied Probability, **Vol. 31**, Studies in Applied Probability (1994), pp. 49-62.

[34] F. Dyson, *Correlations between eigenvalues of a random matrix,* Comm. Math. Phys., **vol. 19** (1970), pp. 235-250.

[35] N. El Karoui, Y. Jiao, *Stein's Method and Zero Bias Transformation for CDOs tranche pricing,* Finance and Stochastics (2008), doi `doi:http://dx.doi.org/10.1007/s00780-008-0084-6.`

[36] C-G. Esséen, *On the Liapunoff limit of error in the theory of probability,* Arkiv för matematik, astronomi och fysik **A28**, 1-19. (1942).

[37] W. Ewens, *The sampling theory of selectively neutral alleles,* Theoretical Population Biology, **vol. 3** (1972), pp. 87-112.

[38] W. Feller, *An introduction to probability theory and its applications,* Princeton University press, Volume I, 3rd edition (1968), Volume II, 2nd edition (1971).

[39] V. Féray, P.-L. Méliot, A. Nikeghbali, *Mod-phi convergence and precise deviations,* Preprint (2013).

[40] R. Fortet, *Sur une suite également répartie,* Studia Mathematica, **Vol. 9** (1940), pp. 54-69.

[41] Y.V. Fyodorov, J.P. Keating, *Freezing transitions and extreme values: Random Matrix Theory, $\zeta(1/2+it)$, and disordered landscapes,* preprint available at `http://arxiv.org/abs/1211.6063`.

[42] Y.V. Fyodorov, G.A. Hiary, J.P. Keating, *Freezing Transition, Characteristic Polynomials of Random Matrices, and the Riemann Zeta-Function,* preprint available at `http://arxiv.org/abs/1202.4713`.

[43] A. Gnedin, G. Olshanski, *q-exchangeability via quasi-invariance,* Ann. Probab., **vol. 38**, n. 6 (2010), pp. 2103-2135.

[44] L. Goldstein, G. Reinert, *Stein's method and the zero bias transformation with application to simple random sampling,* Ann. of Appl. Probability, **vol. 7** (4), pp 935-952 (1997).

[45] L. Goldstein, Y. Rinott, *Multivariate normal approximations by Stein's method and size bias couplings,* J. of Appl. Probability, **vol. 33** (1), pp 1-17 (1996).

[46] M. Gonek, C. P. Hughes, J. Keating, *A hybrid Euler-Hadamard formula for the Riemann Zeta function,* Duke Math. J. **Vol. 136**, Number 3 (2007), pp. 507-549.

[47] A. Grandville, *The anatomy of integers and permutations,* preprint (**2008**), `http://www.dms.umontreal.ca/~andrew/PDF/Anatomy.pdf`

[48] C. P. Hughes, *On the characteristic polynomial of a random unitary matrix and the Riemann zeta function,* University of Bristol, PhD thesis, 2001.

[49] C. P. Hughes, J. Keating, N O'Connell, *On the characteristic polynomial of a random unitary matrix,* Comm. Math. Phys., **vol. 220** (2001), pp. 429-451.

[50] C. P. Hughes, J. Najnudel, A. Nikeghbali, D. Zeindler, *Random permutation matrices under the generalized Ewens measure,* Annals of Applied Probability, **vol. 23**, no. 3 (2013), pp. 987-1024.

[51] C. P. Hughes, A. Nikeghbali, *Zeros of random polynomials cluster uniformly near the unit circle,* Compositio Mathematica, **144** (2008), Part 3, p 734–746.

[52] C. P. Hughes, A. Nikeghbali, M. Yor, *An arithmetic model for the total disorder process,* Probab. Theory and Related Fields, **vol. 141**, no. 1-2 (2008), pp. 47-59.

[53] J. Jacod, E. Kowalski, A. Nikeghbali, *Mod-Gaussian convergence: new limit theorems in probability and number theory,* Forum Mathematicum, **23** (4) (2011), pp. 835–873.

[54] K. Johansson, *On Szegö's asymptotic formula for Toeplitz determinants and generalisations,* Bull. Sc. Math., **vol. 112** (1988), pp. 257-304.

[55] K. Johansson, *On random matrices from the compact classical groups,* Annals of Mathematics, **vol. 153** (2001), pp. 259-296.

[56] K. Johansson, *Discrete orthogonal polynomial ensembles and the Plancherel measure,* Annals of Mathematics, **vol. 153** (2001), pp. 259-296.

[57] A. Jucys, *Symmetric polynomial and the center of the symmetric group ring,* Reports Math. Physics, **t. 5** (1974), pp. 107-112.

[58] M. Kac, *On the distribution of values of sums of the type* $\sum f\left(2^k t\right)$*,* Annals of Mathematics, Second Series, **Vol. 47**, No. 1 (Jan., 1946), pp. 33-49.

[59] M. Kac, *Statistical independence in probability, analysis and number theory,* The mathematical association of America, Wiley and Sons (1959).

[60] N. Katz, P. Sarnak, *Random matrices, Frobenius eigenvalues and monodromy,* AMS Colloquium Publications, **45** (1999).

[61] N.M. Katz, P. Sarnak, *Zeros of zeta functions and symmetry,* Bull. Amer. Math. Soc., **36** (1999).

[62] J.P. Keating, N.C. Snaith, *Random Matrix Theory and* $\zeta(1/2+it)$*,* Comm. Math. Phys. No. 214 (2000), pp. 57-89.

[63] J.P. Keating, N.C. Snaith, *Random matrix theory and L-functions at* $s = 1/2$*,* Comm. Math. Phys., **214** (2000), p 91–110.

[64] S. Kerov, *Asymptotic representation theory of the symmetric group and its applications in analysis,* Translations of Mathematical Monographs, **n. 219** (2003), American Mathematical Society, Providence, RI.

[65] R. Killip, I. Nenciu, *Matrix models for circular ensemble,* Intern. Math. Res. Not., **50**, (2004), pp. 2665-2701.

[66] J. F. C. Kingman, *The representation of partition structures,* J. London Math. Soc., **vol. 18**, no. 2 (1978), pp. 374-380.

[67] E. Kowalski and A. Nikeghbali, *Mod-Gaussian convergence and the value distribution of $\zeta(1/2+it)$ and related quantities,* J. London Math. Soc., **86** (2) (2012), pp. 291-319.

[68] E. Kowalski and A. Nikeghbali, *Mod-Poisson convergence in probability and number theory,* Intern. Math. Res. Not., **18**, pp. 3549-3587, (2010).

[69] C.-A. Laisant, *Sur la numération factorielle, application aux permutations,* Bulletin de la S. M. F., **tome 16** (1888), pp. 176-183.

[70] S. P. Lloyd, *Ordered Prime Divisors of a Random Integer,* The Annals of Probability, **Vol. 12**, No. 4 (Nov. 1984), pp. 1205-1212.

[71] I.G. Macdonald, *Symmetric functions and Hall polynomials,* Oxford Mathematical Monographs, Second edition, The Clarendon Press Oxford University Press (1995)

[72] C. L. Mallows, *Non-null ranking models. I,* Biometrika, **t. 44** (1957), pp. 114-130, `http://www.jstor.org/stable/2333244`.

[73] F. Mezzadri, *How to generate random matrices from the classical compact groups,* Notices of the AMS, **vol. 54** (2007), pp. 592-604.

[74] F. Mezzadri, N.C. Snaith (editors), *Recent Perspectives in Random Matrix Theory and Number Theory,* London Mathematical Society Lecture Note Series (CUP), **322** (2005).

[75] H. L. Montgomery, *The pair correlation of zeros of the zeta function,* Proc. of Symposia in Pure Math., **vol. 24**, American Mathematical Society (1973), pp. 181-193.

[76] G. W. Morgenthaler, *A Central Limit Theorem for Uniformly Bounded Orthonormal Systems,* Transactions of the American Mathematical Society, **Vol. 79**, No. 2 (Jul. 1955), pp. 281-311.

[77] G. Murphy, *A new construction of Young's seminormal representation of the symmetric group,* Journal of Algebra, **t. 69** (1981), pp. 287-291.

[78] Yu. A. Neretin, *Hua type integrals over unitary groups and over projective limits of unitary groups,* Duke Math. J., **vol. 114** (2002), pp. 239-266.

[79] I. Nourdin and G. Peccati, *Stein's method on Wiener chaos,* Probab. Theory Related Fields (2009).

[80] G. Olshanski, *Random permutations and related topics,* The Oxford Handbook of Random Matrix Theory, **ch. 25**, G. Akemann, J. Baik, P. Di Francesco, eds (2011).

[81] P.-O. Dehaye, *Joint moments of derivatives of characteristic polynomials,* Algebra and Number Theory, **2** (1), (2008), p 31–68.

[82] E. M. Rains, *High powers of random elements of compact Lie groups,* Probability Theory and Related Fields, **vol. 107**, no. 2 (1997), pp. 219-241.

[83] A. Rényi, P. Túran , *On a theorem of Erdös-Kac,* Acta Arithmetica, **vol 4**, pp. 71-84 (1958).

[84] V. Rotar, *Stein's method, Edgeworth's expansions and a formula of Barbour,* (2004)

[85] N. Ross, *Fundamentals of Stein's method,* preprint (2011), available at `http://arxiv.org/abs/1109.1880`

[86] Z. Rudnick, P. Sarnak *Zeros of principal L-functions and random matrix theory,* Duke Math. J., **81** (1996), p 269–322.

[87] A. Selberg, *Bemerkninger om et multipelt integral,* Norsk. Mat. Tidsskr., **vol. 24**, no. 2 (1944), pp. 71-78.

[88] A. Selberg, *Note on a paper by L.G. Sathe,* J. Indian Math. Soc., **vol 18** pp. 83-87 (1954).

[89] A. Selberg, *Old and new conjectures and results about a class of Dirichlet series,* in Proceedings of the Amalfi Conference on Analytic Number Theory, September 1989, (1992), pp. 367-385.

[90] N.C. Snaith, *Riemann zeros and random matrix theory,* Milan Journal of Mathematics, **78** nb. 1 (2010), pp. 135-152.

[91] F. Spitzer, *Some theorems concerning 2-dimensional Brownian motion,* Transactions of the American Mathematical Society, **Vol. 87**, No. 1 (1958), pp. 187-197.

[92] S. Starr, *Thermodynamic Limit for the Mallows Model on S_n,* J. Math. Phys., **Vol. 50** (2009).

[93] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables,* Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, pp. 583-602 (1972).

[94] C. Stein, *Approximate computations of expectations,* Institute of Mathematical Statistics, Lecture Notes-Monograph Series, **vol. 7** (1986).

[95] E. Strahov, *Moments of Characteristic Polynomials Enumerate Two-Rowed Lexicographic Arrays,* Electronic Journal of Combinatorics, **vol. 10** (2003).

[96] G. Szegö, *Orthogonal Polynomials,* Publications of the AMS, 1955 (1st ed. 1939).

[97] T. Tao, *Structure and randomness, t. 2 , Selberg's limit theorem for the Riemann zeta function on the critical line,* available at `http://terrytao.wordpress.com/2009/07/12/selbergs-limit-theorem-for-the-riemann-zeta-function-on-the-critical-line`.

[98] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory,* Cambridge studies in advanced mathematics, **vol. 46**, Cambridge University Press (1995).

[99] E.E.C. Titchmarsh, *The Theory of the Riemann Zeta-Function,* Oxford University Press (1986).

[100] C. Villani, *Optimal transport, Old and new,* Springer-Verlag (2009).

[101] G. A. Watterson, *The stationary distribution of the infinitely-many neutral alleles diffusion model,* J. Appl. Prob, **vol. 13** (1976), pp. 639-651.

[102] G. Murphy, *The classical groups, their invariants and representations,* Princeton University Press, Princeton, N.J. (1939).

[103] K. Wieand, *Eigenvalue distributions of random unitary matrices,* Probab. Theory Relat. Fields, **vol. 123** (2002), pp. 202-224.

[104] K. Wieand, *Eigenvalue distributions of random permutation matrices,* Ann. of Probab., **vol. 28** (2000), pp. 1563-1587.

[105] A. Zygmund, *Trigonometric series,* Volume II, Cambridge (1959).